HEC MONTRÉAL

Emotion perceived in high school students in relation to their lifestyle habits

par

Jessica Leonard

Pierre-Majorique Léger
HEC Montréal
Codirecteur de recherche

Marc Fredette
HEC Montréal
Codirecteur de recherche

Sciences de la gestion
(Spécialisation Intelligence d'affaires)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

December 2022

Résumé

En collaboration avec Dis-Moi, une organisation à but non lucratif, nous avons effectué une enquête longitudinale sur l'auto-évaluation de la santé mentale des élèves sur une base volontaire ou en classe d'août 2020 à mai 2021 pendant 20 semaines. L'objectif de cette étude est d'aider les éducateurs scolaires à mieux comprendre leurs élèves et à leur prodiguer des conseils et des outils appropriés en se basant sur les réponses d'enquêtes.

En utilisant le regroupement de courbes sur leurs émotions, nous avons identifié des groupes qui se différencient sur non seulement le sexe et l'année scolaire, mais aussi le style de vie et le niveau émotionnel auxquels ces étudiants s'identifient. Nos résultats suggèrent trois groupes différents: l'étudiant moyen, l'étudiant fatigué et l'étudiant énergique. Les deux variables significatives influençant les étudiants dans leurs groupes respectifs sont le niveau d'énergie perçu par l'élève et le niveau de motivation à l'école. Nous concluons que le groupe des étudiants énergiques présente un taux d'énergie élevé et une motivation élevée à l'école, s'en sort mieux que tout autres groupes et semble beaucoup plus heureux. À contrario, les individus du groupe « étudiants fatigués » indiquait être moins heureux et plus anxieux.

Nous pouvons apporter des suggestions pour augmenter le bien-être des étudiants s'ils sont placés dans le regroupement des "étudiants fatigués", les agents de santé devraient se concentrer sur leur niveau d'énergie et de motivation à l'école pour les aider à combattre leur faible niveau d'émotions. Nous pouvons également suggérer que même si les élèves moyens ne sont pas à risque, ils devraient tout de même être suivis et bénéficier d'un certain soutien pour les aider à réussir et à être heureux en cours de route. Cela démontre que les conseillers scolaires et les enseignants sont tout aussi importants que les parents pour promouvoir la santé mentale et que des technologies comme les agents virtuels et les activités en classe sont disponibles pour encourager ce discours. Cet article ajoutera à la littérature, car il s'agit de la première étude connue utilisant le regroupement de courbes dans le temps dans un contexte de santé mentale.

Mots-clés : regroupement de courbes, santé mentale, école secondaire, adolescents, clusters, émotions, régression multinomiale, comparaison en paire, matrice de corrélation.

Abstract

In collaboration with *Dis-Moi*, a non-profit organization, we collected a longitudinal survey on self-assessed students' mental health on either a voluntary or in-class basis from August 2020 to May 2021 for 20 weeks. The goal of this study is to help school educators understand their students better and give them proper guidance and tools adapted to the answers on the surveys.

By using curve clustering on their emotions, we found clusters that can help differentiate not only by gender and school year but by lifestyle and emotional level these students are facing. Our results reveal three different clusters: the average student, the fatigued student, and the energetic student. The two significant variables influencing the students to fall into their respective clusters are the level of energy and level of motivation at school. We conclude that the cluster with the energetic students with high energy and high motivation at school do better off than any other cluster and were much happier. Vice versa, the fatigued students were less happy and more anxious.

We can bring suggestions to increase students' well-being if they are placed in the "fatigued student" cluster, health workers should focus on their level of energy and motivation in school to help them combat their low level of emotions. We can also suggest that even if the average students aren't at risk, they should still be monitored and given some support to help them succeed and gain happiness along the way. This demonstrates that school counselors and teachers are just as important as parents to promote mental health and technologies like virtual agents and classroom activities are available to encourage this speech. This paper will add to the literature as it is the first known study using curve clustering over time in a mental health setting.

Keywords: curve clustering, mental health, high school, teenagers, clusters, emotions, multinomial regression, pair-wise comparison, curve clustering, correlation matrix, mean analysis

Table of Contents

List of tables and figures

Preface

This thesis was written by Jessica Leonard for the M.Sc. program at HEC Montréal. The research project was approved by the Research Ethics Committee (REC) of HEC Montréal in March 2021. HEC Montreal authorized this paper to be written in English.

Acknowledgements

I would first like to thank my co-directors, Pierre-Majorique Léger and Marc Fredette, for their exceptional mentorship during my research project. It is because of their patience, their advice, and their feedback that I was able to complete this project.

Special thanks to Shang Lin for his help in the analysis throughout the project. His attention to detail and his assistance helped me to carry out each analysis in a diligent and applied way.

Finally, I would like to thank my family and my friends for their unconditional support throughout my master's thesis. They were my moral support throughout this process and couldn't have done it without them.

Chapter 1: Introduction

Problematic

The emergence of the coronavirus (Covid-19), in December 2019, had the world in a state of emergency crisis. Hospitals were over-capacitated, grocery stores were dealing with food shortages, offices were closing, workers were sent to work from home, and curfews were imposed. In over 190 countries, roughly 1.6 billion students were impacted by school closures and had to pivot to online learning (Akmal et al., 2020). This crisis caused a high level of stress and anxiety in both adults and children as the "new" normal imposed by the government broke all sense of routine and in-person social interactions. Distress was brought to many students which resulted in a significant increase in mental health issues, illness perception, anxiety, and depression symptoms (Aqeel, 2021).

Studies have demonstrated that epidemics, diseases that affect people within a community, a population, or a region, have a significant impact on the public's mental health (Liang, 2020). Diseases such as SARS and H1N1 lead to issues such as post-traumatic stress disorder, psychological distress, depression, and anxiety in certain groups (Liang, 2020). A study on the effect of Covid-19 on youth mental health can suggest that the pandemic affected nearly 40.4% of the youth as they were now more vulnerable to psychological problems and nearly 14.4% in this sample had post-traumatic stress disorder (PTSD) symptoms (Liang, 2020). The lack of longitudinal data on change over time throughout the pandemic limits our understanding of how certain components of the pandemic, like the national lockdown (which included school closures for most children), were associated with changes in mental health (Waite et al., 2021).

Dis-Moi, a non-profit organization, decided to make a difference by taking the initiative to remotely monitor mental health in high schools. Their objective was to increase students' well-being, demystify mental health and reduce students' psychological distress. They created an online longitudinal survey that was sent out once a week asking students about different topics regarding their overall health. These are the targeted topics: how they are feeling today, their level of physical activity, their average level of sleep each night, their average level of energy, their level of stress/anxiety, their level of relationship with their parents, their level of motivation in school, the amount of time they spend in front of a screen excluding schoolwork and their level of communication with their family.

This study is important as it tracks over time the mental health of teenagers. The lack of longitudinal data on change over time associated with changes in mental health is a gap we'd like to resolve with this study. Our proposed solution is to use curve clustering to understand these students and find out if we can cluster them into groups. In doing so, we will be able to pinpoint differences in these students' lifestyles, and counselors will have more information to understand how these different factors affect their lives. We know that differences in measurement may contribute to discrepancies in the research about changes in the prevalence of mental illness in the general population (McMartin et al., 2014). In the same ways, a health professional's diagnosis of a problem determines the findings of certain research, whereas parent or teacher reports of symptoms or in-person interviews determine the results of other studies (McMartin et al., 2014). Our study will be self-assessed by the students themselves as they will be answering a survey and it will be followed over the nine months of the study.

It is currently difficult for healthcare professionals to detect which types of students need mental health support and how to help once identified. As it was explained, curve clustering to segment groups of students based on their longitudinal trends will allow them to easily categorize and help those in need. Our findings will have impacts on how we divide teenagers' mental health, how counselors should look at the issue, and how research should be conducted. Further research can be done using our method on a broader number of students to achieve more valuable results.

Research Question

The objective of this study is to demonstrate that we can segment groups of students over time by their emotions to detect which types of students need mental health support and how to help once identified.

Our research question is as follows:

Is it possible to find clusters of students over time based on the longitudinal similarities/differences in their emotions and find characteristics that differentiate groups?

Our hypothesis:

H0: With curve clustering analysis techniques over time, we anticipate finding different meaningful clusters of students.

We will analyze to find that the level of energy and motivation are the main characteristics explaining these groups. The remainder of this paper is organized as follows. A literature review on mental health issues amongst teenagers, methods for tracking mental health over time, and our hypothesis development based on curve clustering, will be introduced. Afterward, our methodology will be described as such: the study design, the research participants, the measures, and the analysis. The results will be organized by our clustering technique and the resulting clusters found followed by a conclusion where a return on the results, research limitations, and future research will be discussed.

Chapter 2: Literature Review

This current section aims to first paint a portrait of mental health amongst teenagers, then a review of solutions to teenagers' mental health issues by evaluating both the intervention methods proposed by front-line workers and mental health organizations and how curve clustering is used in our case to establish our findings. After this, a hypothesis will be given based on the research question as to how we will tackle these issues in our paper.

1. Mental health issues amongst teenagers

Although we are collectively aware of mental health issues, rarely do we dive in to truly understand the state of the problem. As organizations and health institutions try to answer our society's need for better and more accessible mental health services, our study proposes to go upstream and analyze how we address mental health issue early on in an individual's life. This subsection sets the table for the purpose of our study by establishing why more research is needed on teenagers' mental health and why high schools represent the ideal social structure to further positive mental health amongst our future citizens.

1.1 High Schools: a steppingstone for a mentally healthy society

As indicated by the World Health Organization (WHO), given that "adolescents comprise 16% of the global population, it is vital to address the main threat to their health in order to achieve SDG (Sustainable Development Goals) targets"(Organization, 2020). The literature recognizes that adolescence represents a sensitive period of life, shaping an individual's ability to navigate social situations and hardships as an adult, hence the importance to provide adequate mental health services and education to this population (Blakemore & Mills, 2014). This can be due to physical changes (i.e., neurological changes) and the impact of the social environment on a teenager's socialization (Blakemore & Mills, 2014). Neurological changes may impact a teenager's response to a change in their social environment, but the environment has a considerable effect on a teenager's development (Benningfield et al., 2015). More precisely, as the pre-frontal cortex changes, teenagers may be "[…] more affected by distress" (Benningfield et al., 2015) and they may "[…] be more vulnerable to psychiatric illnesses" (Blakemore & Mills, 2014). Effectively, the WHO indicated that half of the mental health conditions emerge before the age of fourteen years old (Organization, 2020). Combined with the physiological change occurring, emerging stressors and challenges occurring during this period of life significantly impacts adolescents' likelihood to develop mental health issues (Blakemore & Mills, 2014). Puberty comes with integration into high school and it implies changes in the environment, new peers, and new learning strategies needed to be developed (Blakemore & Mills, 2014). Similarly, as social connections tend to change, it also impacts teenagers aged thirteen years old more significantly as they start seeking social acceptance and fearing social rejection (Blakemore & Mills, 2014).

Most of the changes described above take place in a learning institution. Therefore, high schools represent the ideal social structure through which mental health-related issues can be identified and addressed, and where mental health literacy among teenagers can be furthered (Kutcher et al., 2015). Because early detection and intervention are critical, high schools position themselves as the front line of the battle against mental health issues, with teachers and intervenors as the first resource available (Organization, 2020). As reported by the World Health Organization in 2020, universally delivered interventions aimed at promoting positive mental health could have a positive impact on "[…] suicidal behaviors, mental disorders (such as depression and anxiety), aggressive-disruptive and oppositional behaviors, and substance use" (Organization, 2020).

Universal programs do not need to target specifically at-risk persons and can be offered school-wide or grade-specific to adjust to the unique needs of the student's age and school year (Badawi et al., 2022). Given that literacy on diverse subjects is though in school, adding mental health literacy would be both "reasonable and rational" (Kutcher et al., 2015). The research established that the cognitive process required to perform academically is closely linked to social and emotional health, making it even more obvious that schools' mission should include both furthering learning opportunities in traditional subjects (i.e. maths, language classes, etc.) and in mental health classes (i.e. emotional regulation, mindfulness, seeking help, etc.) (Organization, 2021). As we now understand how critical it is to intervene in teenagers' mental health education and that high school presents itself as the best stage to do so, we turn to evaluate the state of the situation.

1.2 Current situations regarding mental health in teenagers

Interestingly, up-to-date statistics regarding our age group (11-16) are limited, indicating that further studies are warranted. Reports consider “youth” as individuals aged between fifteen to thirty years old or aged from twelve to seventeen years old (Canada, 2019). Therefore, in the literature, eleven years old individuals are left out of the statistics as they tend to be considered children still.

As data sources and studies on this subject are concerned, we find three main reliable sources: the Health Characteristic data collected through the *Canadian Community Health Survey (CCHS)* – most recent data being up until 2021 (Canada, 2019), the *Enquête Québécoise des Jeunes du Secondaire (ISQ)* – most recent report done in 2016-2017 (Street, 2018), and the *Portrait du Bien-être des jeunes au Québec* – most recent report done in 2019 (Gallant et al., 2019) and heavily based on the CCHS results (gathered between 2007-2016) and the *ISQ* report. The ISQ report is of particular interest for this study, as it gathers information specifically on individuals in high schools and differentiates the results by school levels. The limitation of the *ISQ* report is its location as it only targets students in Quebec, therefore being a smaller sample than the CCHS. Meanwhile, the “Portrait of Youth in Canada: Data Report – Chapter 1: health youth in Canada” could have been interesting, but the age group considered (15-30 years old) is too broad for the context of this study. Although these reports are excellent starting points, it also underlines the need for more data collection regarding mental health among individuals in high schools.

Based on the reports, we could divide indicators into two broad categories: the subjective and the objective ones. The subjective indicators are self-declared by the respondent and are reflective of one’s perception of their mental state. The subjective indicators are “perceived mental health”, “life satisfaction” and “perceived life stress. In counterpart, the objective indicators are diagnosed mental health issues and are not subject to the opinion of the respondent with indicators such as “mood disorders”, “mental health issues”, “use of drugs prescribed for anxiety, depression, for calmness or for concentration” (Street, 2018).

We now turn to subjective indicators. Perception of the respondents on the “perceived mental health” axis is tracked through the status of the indicator being either “very good or excellent” or “fair or poor” (Canada, 2019). This indicator allows us to zoom in on the issue of mental health, which might not be captured in the more general “perceived health” indicator (Canada, 2019). While 81.5% of the respondents indicated having very good or excellent mental health in 2007-2018 (Gallant et al., 2019), this number has dropped by 19.2% since then (see Figure 2) (Canada, 2019). Alarmingly, respondents perceiving their mental health as “poor or fair” grew by 7.8 percent between 2015-2021 (see Figure 3) (Canada, 2019). The change in mental health statistics becomes more alarming when differentiated by sex. Based on the statistics from the *CCHS* females’ mental health has been declining more than their male counterparts. The percentage of females perceiving their mental health as “very good or excellent” has decreased by 23.4% between 2015-2021 while males’ situation has decreased by 8.1% (see Figure 2). Concurrently, there’s been an increase of 11% in females declaring that their mental health was “fair or poor” and an increase of 4.6% for males for the same reference period (see Figure 3). We also observe a more drastic change in mental health around 2019 which coincides with the beginning of the COVID-19 pandemic. The impact of the pandemic on mental health will be later discussed. A similar indicator is psychological distress, as measured in the ISQ survey, which was measured with questions asking the respondents to self-report symptoms related to diverse mental health issues (Street, 2018). The conclusion is the same: girls presented more signs of psychological distress (Street, 2018). In addition, students in Secondary 4 and 5 were experiencing higher levels of psychological distress (35%) than students in lower levels (21%-30%) (Street, 2018).

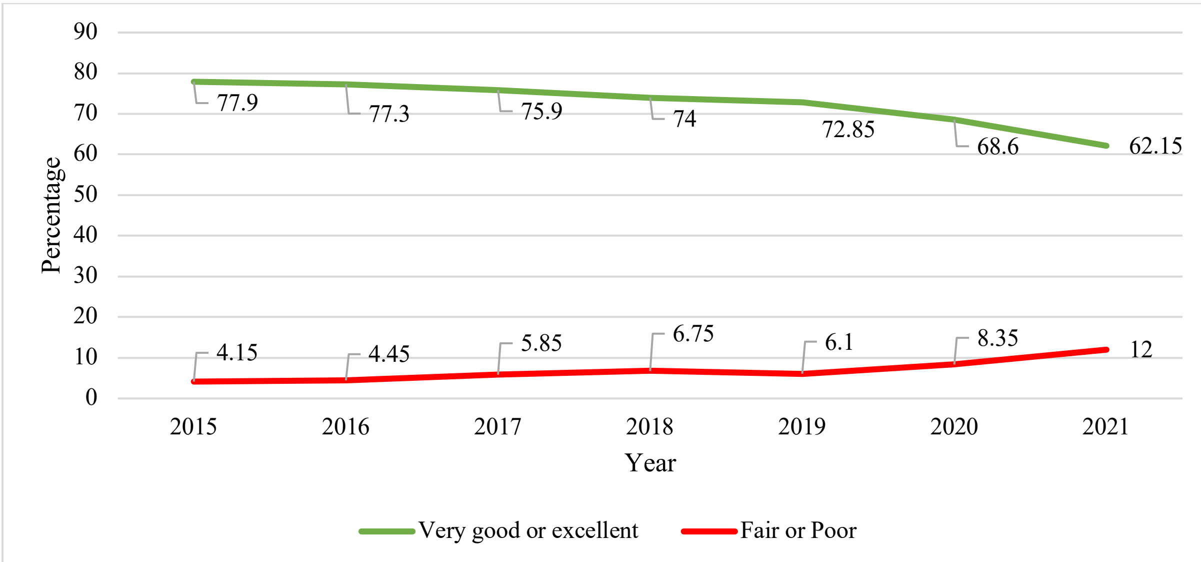

Figure 1: Perceived mental health (%) in respondents aged 12-17 years old, both sexes, from 2015-2021(Canada, 2019)

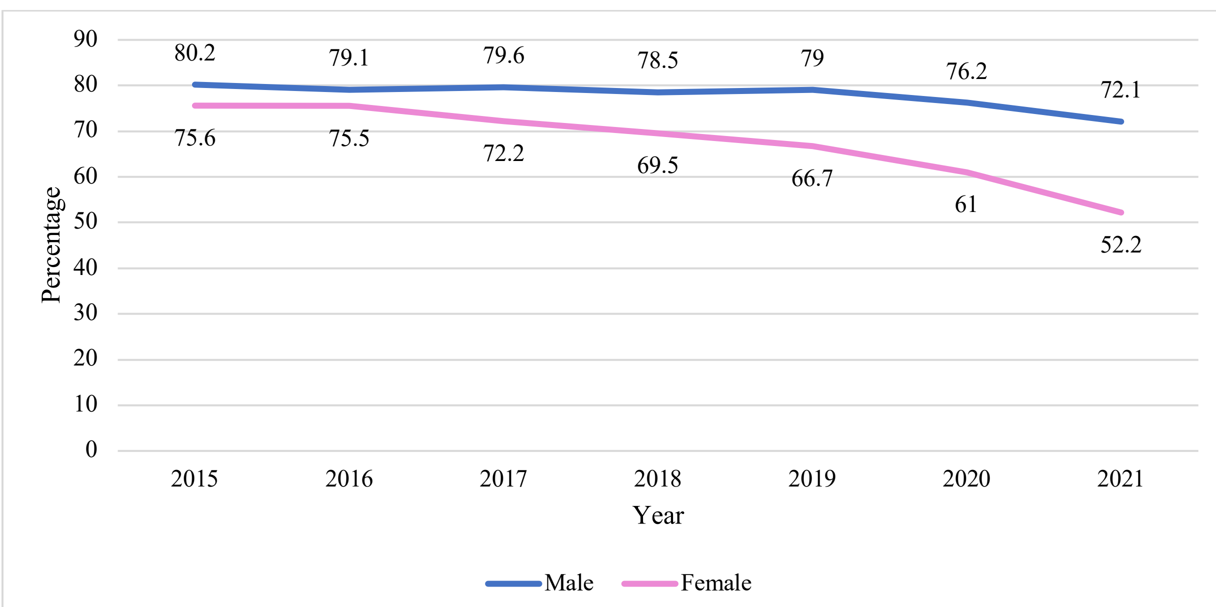

Figure 2: Perceived health as very good or excellent for respondents between 12 to 17 years old (%) differentiated on sex, between 2015-2021 in Canada (Canada, 2019)

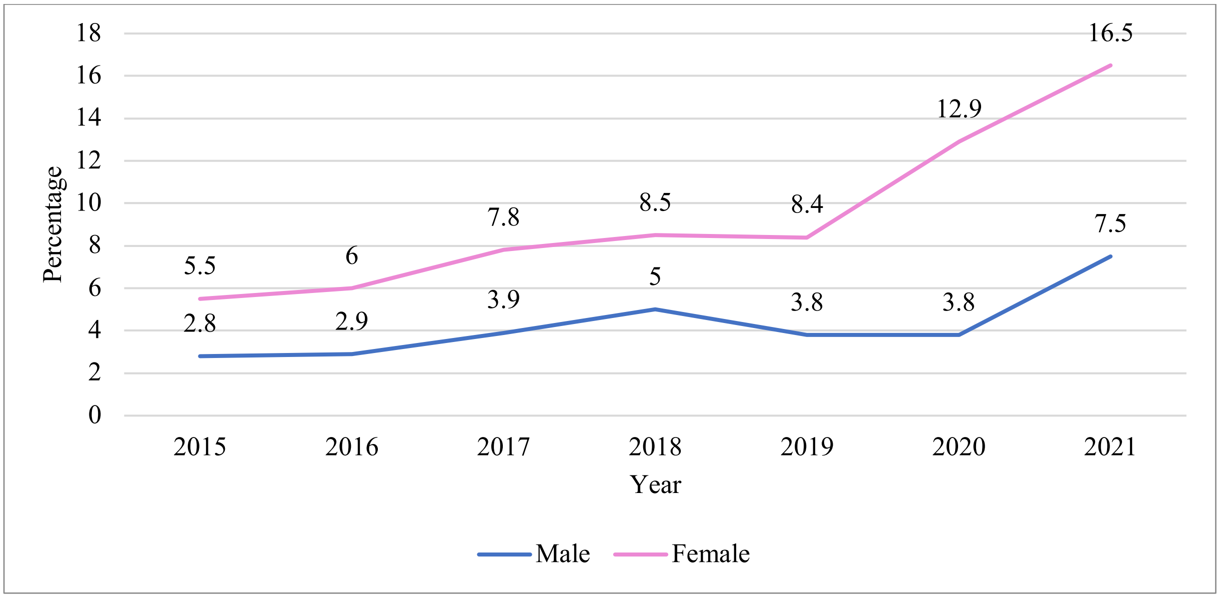

Figure 3: Perceived mental health as fair or poor for respondents between 12 to 17 years old (%) differentiated on sex, between 2015-2021 in Canada (Canada, 2019)

The second subjective indicator is life satisfaction which represents individuals declaring they were satisfied or very satisfied with their life (Canada, 2019). Until 2019, life satisfaction had been relatively stable, with boys declaring a higher life satisfaction than girls (see Figure 4)**.** After that year, girls' life satisfaction dropped by 4.9% whereas boys decreased by only 0.7% and even increased by 0.1% from 2021-2022. The term "perceived life stress" refers to the level of stress a person perceives in their life on most days (Canada, 2019). Girls declare significantly higher life stress (around 15%-20%) than boys (8.7% - 9.5%) between 2015-2021 (see Figure 5). To resume the "subjective indicators" section, we can conclude that girls' self-perceived mental health is lower than boys.

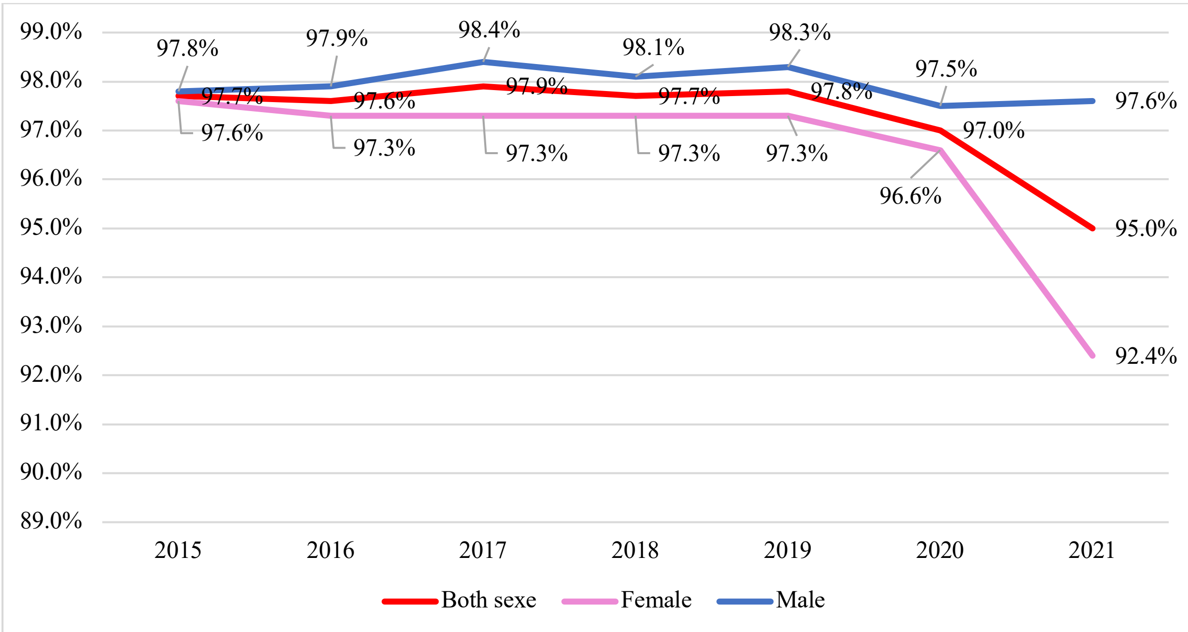

Figure 4: Declared life satisfaction by respondents aged between 12-17 years old, from 2015-2017

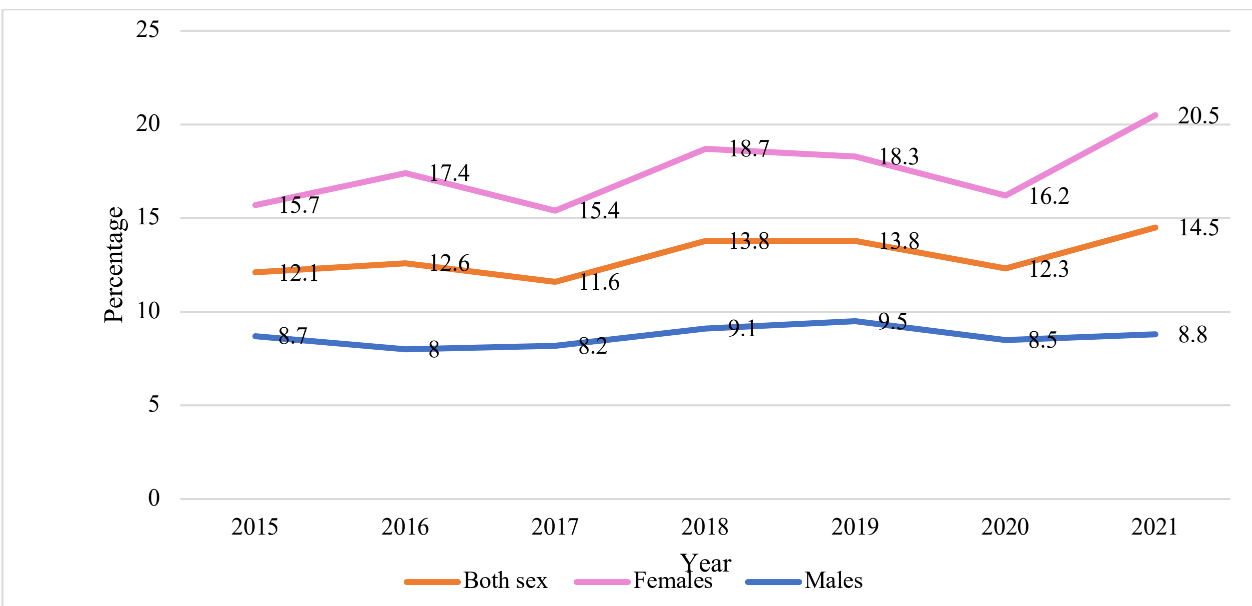

Figure 5: Perceived life stress as "most days quite a bit or extremely stressful" by respondents aged between 12-17 years old, between 2015-2021

In terms of objective indicators, the "mood disorder" indication denotes those who have been diagnosed by a health professional with a mood disorder, such as depression or bipolar disorder

(Canada, 2019). Mood disorders were slowly going down from 2015-2019 but saw an increase between 2019-2021 (see Figure 6) (Canada, 2019). From 2019-2020 there was a decrease of 4.3%, the lowest point reached since 2015. Hypothetically, this could be due to the reduced access to mental health practitioners and the health system during the pandemic, thus decreasing the number of official diagnoses. In 2021, we saw the largest increase as it went from 4.3% to 5.6%. The number of official diagnoses contrasts drastically with the self-reported indicators, which may indicate that although teenagers are feeling distressed, they struggle to get the proper resources, and/or the system struggles to identify the student in distress. The ISQ measures a similar indicator, "Mental health Issues" and presents more precise results. 17% of the respondents had been diagnosed with anxiety disorders, 6% were diagnosed with depression and 2.2% were diagnosed with an eating disorder (Street, 2018). Aggregated, these mental health issues were present in 20% of the students (Street, 2018). There's a striking difference between the results of the *ISQ* and the *CCHS*. Although these studies were done at a different scale, we must wonder if this difference may be correlated positively with the fact that the ISQ was performed in a school setting. This would align with the literature, indicating that schools are better placed as first interveners for mental health issues. Based on the ISQ, students grow more anxious, tend to be more depressive, and are more likely to have eating disorders as they advance at different levels (Street, 2018). As for learning difficulties (ADHD), the prevalence remains similar in all levels of high school (Street, 2018). Such data coincides with the growing need for acceptance by peers and the important life changes that occur as students become older.

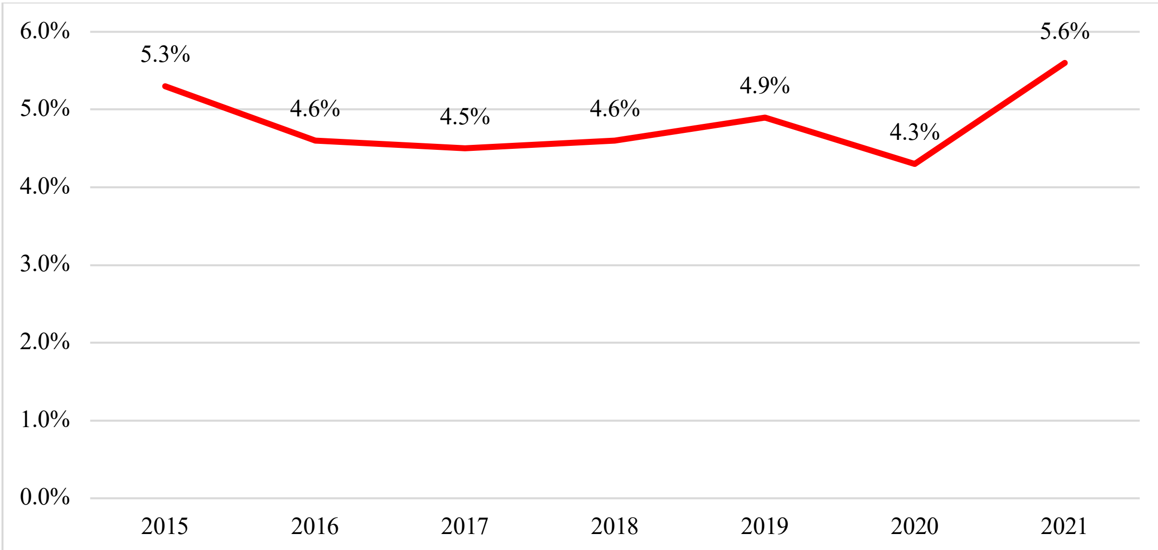

Figure 6: Mood disorder for 12 to 17 years for both sexes (Canada, 2019)

The last objective indicator, only available in the ISQ report, regards the use of prescribed drugs. Students use drugs more to calm themselves or to focus (14.8%) than to treat anxiety and depression (3.6%) (Street, 2018). There was no significant difference between levels of anxiety and depression drugs, but there was for the use of drugs for calmness and concentration (Street, 2018). To resume the objective portion of the indicators, it appears that although they speak for a worsening situation, they do not reflect the self-reported mental health of teenagers. Hence, the importance of ensuring that we better our detection of mental health struggles by potentially using schools to do so.

Now that we understand better the state of the situation, we must look at ways for front-line workers in schools use to help students combat their mental health issues. However, we understand that these subjects could be the subject of an entire paper and therefore do not pretend to be exhaustive. It is rather a more superficial exploration of the literature to provide a better context for our results.

2. Intervention Methods

Teachers frequently report seeking assistance in creating wellness in their classrooms as well as recognizing and helping students with mental health issues (Organization, 2021). It is quite common that a few misfits disrupt the classroom, and the solution is detention, but what if they needed better care? Managing challenging classroom behavior and fostering students' academic performance are both made simpler for instructors when they feel more prepared to help students' mental health (Organization, 2021). That's why in this chapter, we will be reviewing why school counselors and teachers are so important, how to promote mental health, what studies have revealed about teenagers' mental health over time, and the technologies that help deal with these issues.

2.1 Importance of investing in mental health in schools

School counseling has been around since the early 1900s as vocational guidance (Gysbers, 2012). It was formed in schools as a role filled by teachers and administrators (Gysbers, 2012). In the 1920s, the movements for child study, psychometric testing, and mental health all had an impact on school counseling and a more clinically based school counseling approach arose (Gysbers, 2012). This marked a change from economic to psychological difficulties, with a focus on therapy for personal adjustment (Gysbers, 2012).

Students spend about 15 000 hours from kindergarten through graduation and can benefit from mental health promotion, prevention, and intervention (Rutter, 1991). Investing in mental health has many beneficial impacts on academic success rates while lowering dropout and grade retention rates (Organization, 2021). Studies have provided evidence that the degree of school attachment, engagement, and commitment, is highly correlated with more favorable academic outcomes (Stewart, 2008). School programs that focus on mental health are linked to students' academic accomplishment (Gutman & Feinstein, 2008), academic success and motivation may be improved via social development, which includes meaningful peer interactions that boost psychological well-being and life skills (Stewart, 2008), and promoting mental health in schools contribute to a drop in violence and youth criminality (Organization, 2021).

For teachers, working with kids who have mental health issues may be extremely difficult and frustrating (Hanko, 2018). Improving students’ emotional well-being can boost instructors' retention and satisfaction (Hanko, 2018).

All in all, schools that encourage the adoption of good behaviors such as healthy diets, physical exercise, etc., rather than addictions, delinquency, etc., will increase students’ overall quality of life (Organization, 2021). There is a multitude of benefits that come with properly implemented school counseling services and these services are here to stay for the foreseeable future (Gysbers, 2012).

2.2 Promoting mental health in every setting

Promoting mental health development in schools is the first step any institution should take to raise awareness. This method is the most effective as it teaches the principles of how to build a strong character and how to keep up a good lifestyle. The Mental Health Foundation has listed essential characteristics of schools that support students' mental health, including having a senior leadership team dedicated to fostering an environment where each child is valued and respected regardless of their abilities, valuing teachers, non-teaching staff, and everyone involved in the care or supervision of students, and having school-wide policies on crucial issues like behavior and bullying that are clearly laid out, accepted, and implemented throughout the school (Kay, 1999).

Values are also very important to promote at a young age to understand right from wrong. The World Health Organization enumerated a few that a school should instill in their learning program to cater to the students such as caring for all, valuing diversity, building self-esteem, building relationships, ensuring safety, encouraging participation, early identification and intervention to promote well-being and mental health, and support and training for teachers and other staff (Organization, 2021). Teachers are not expected to be qualified counselors, they can employ methods like circle time also known as “group thinking time” or positive discipline techniques to ensure compliance, but if required, schools should direct kids who require counseling to professionals (Organization, 2021).

Parents have an important part in their children's education both in and out of the classroom. Parental involvement is vital for both boys’ and girls' school success, but an active parent has a considerably greater influence on girls' overall school achievement than boys (Stevenson & Baker, 1987). Ways to keep a healthy environment at home is by supporting the child’s education at school and at home (i.e., prioritizing school, being a good role model towards learning, involvement, attendance, etc..), providing an environment that is conducive to learning at home (keeping it calm and quiet, encouraging reading and doing homework, ensuring proper diet and encouraging physical exercise, limiting screen time, etc.), communicating with the school and ensuring their child’s academic attainment, emotional well-being, and social development and maintaining open lines of communication with schools and parents to ensure their children's academic success, emotional stability, and social development (Organization, 2021).

Proper diet is a further non-academic strategy to support young people's mental health (Organization, 2021). It is crucial for maintaining both a healthy body and mind. Children's cognitive and emotional development is impacted by nutritional deficits (Organization, 2021).

The CDC recommends 60 minutes of exercise a day to increase stamina and endurance, supports the development of strong bones and muscles, enhance blood circulation, and aid in weight management (Services, 2022). Additionally, it has positive impacts on mental health, including lowering stress and anxiety levels, preserving a positive body image, and boosting self-esteem (Services, 2022). It also assists in keeping kids busy with constructive activities (Services, 2022).

The American Academy of Pediatrics suggests that parents supervise their children's "media time" (Organization, 2021). Similarly, educators have a similar role to play in growing access to technology in schools and after-school programs (Organization, 2021). As a result, it is critical for parents to be active participants in their children's education alongside the school (Organization, 2021).

2.3 Tracking over time the mental health of teenagers

Existing studies and media reporting give contrasting views of mental health trends. The media tends to report a higher amount of mental health in teenagers, but the studies don't always back those facts. The following studies will be on both sides of the spectrum. Of course, methods change from study to study, so it is quite difficult to follow the same trend, but this shows how important it is to continue the research in this field to find more conclusive results. The first three study show there are little differences in mental health in teenagers compared to the next five studies that show a decline in mental health.

A Canadian study, from 1994 to 2009, investigated over 30 thousand participants aged between 10 to 15 years old (McMartin et al., 2014). Participants completed self-reported assessments of mental disease indicators such as conduct disorder, hyperactivity, indirect aggression, suicidal behavior, and sadness and anxiety every two years (McMartin et al., 2014). Linear regression was used to examine trends in mean scores across time to observe the prevalence of mental illness in these Canadian children and adolescents (McMartin et al., 2014). For participants of all ages, the distribution of scores on depression, anxiety, conduct, and indirect aggression remained consistent or showed minor declines over time (McMartin et al., 2014). Except for hyperactivity, the incidence of mental disease symptoms among Canadian children and adolescents has been largely steady (McMartin et al., 2014). Disagreements in claims of the rising incidence of mental illness in Canada may be explained by differences in study methods, an increase in treatment-seeking behavior, or changes in diagnostic criteria or procedures (McMartin et al., 2014).

Another Canadian study used information from the Canadian Community Health Survey (2001, 2003, 2005, 2007), as well as data from the National Population Health Survey (1994 to 2008) (Simpson et al., 2012). Measures were taken for major depressive episode prevalence, distress, clinically diagnosed mood disorders, antidepressant usage, self-rated subjective mental health, and self-rated stress (Simpson et al., 2012). The results were that depression and severe distress remained stable throughout time (Simpson et al., 2012). However, more people were being diagnosed with mood disorders, and a larger percentage of people said they were taking antidepressants (Simpson et al., 2012). Although fewer people said their lives were highly stressful, the number of those who said they had poor mental health remained constant (Simpson et al., 2012). This shift is most likely the result of shifts in diagnostic practice, mental health knowledge, or willingness to disclose mental health issues (Simpson et al., 2012). However, there was no concrete evidence of a change in mental health status (Simpson et al., 2012).

Another study was comprised of over 900 thousand teenagers from 36 different European countries in the years 2002, 2006, 2010, 2014, and 2018 (Cosma et al., 2020). Cross-national changes in teenage mental well-being and homework pressure were evaluated using hierarchical multilevel models (Cosma et al., 2020). The results showed that stress-related symptoms and academic pressure showed a slight linear rise over time (Cosma et al., 2020). There was no difference in life satisfaction (Cosma et al., 2020). Furthermore, higher-income nations saw decreases in well-being and a rise in academic pressure (Cosma et al., 2020). This study does not support significant decreases in mental well-being among teenagers (Cosma et al., 2020). The slight rise in homework pressure over time explained a portion of the increase in psychosomatic health symptoms across countries (Cosma et al., 2020).

On the other hand, in a UK study in the years 1974, 1986, and 1999, parents of 15–16-year-old filled out comparable surveys (Collishaw et al., 2004). The findings revealed a marked rise in teenage conduct problems of around 1.5 between each cohort which impacts all sexes, all socioeconomic classes, and all kinds of families (Collishaw et al., 2004). There was evidence of a recent increase in emotional issues, however conflicting findings about rates of hyperactive behavior (Collishaw et al., 2004). Hyperactivity was stable for girls during the entirety of the study whereas the boys decreased between the first two cohorts and increase afterward (Collishaw et al., 2004). An analysis of larger socioeconomic patterns influencing teenagers' lives appears likely to give crucial hints as to probable explanations for these mental health trends (Collishaw et al., 2004).

Another interesting case was data from computerized medical records of 42–82 Dutch general practices from patients aged 0–18 (N range from 37716 to 73432) between 2004 and 2008 (Zwaanswijk et al., 2011). Historical trends in the prevalence of documented mental health issues, psychotropic drug prescriptions, and referrals to primary and secondary mental health treatment were observed (Zwaanswijk et al., 2011). Children and teenagers diagnosed with mental health disorders and those who were prescribed psychostimulants have increased whereas antidepressant prescriptions dropped with time in both age groups (Zwaanswijk et al., 2011). Children transferred to primary and secondary mental health treatment grew with time, while teenagers showed no significant increase (Zwaanswijk et al., 2011). Despite an improvement in general practice diagnosis of mental health disorders and referrals to primary mental health treatment, most referrals are still directed to secondary care (Zwaanswijk et al., 2011). Early detection and treatment are critical to avoiding issues that may be more difficult to address in adulthood (Zwaanswijk et al., 2011).

In England, important surveys of children and teenagers' mental health were conducted in 1999, 2004, and 2017 (Sadler et al., 2018). The survey incorporates feedback from instructors, parents, and students (depending on the age of the selected child) on disorders which can be emotional disorders, behavior or conduct disorder, hyperactivity disorder, or other disorders such as autism, eating disorders, tic, etc. (Sadler et al., 2018). Data from this survey series show a small increase in the prevalence of mental disorders among children aged 5 to 15 from 9.7% in 1999 to 10.1% in 2004, it has risen to 11.2% in 2017 (Sadler et al., 2018).

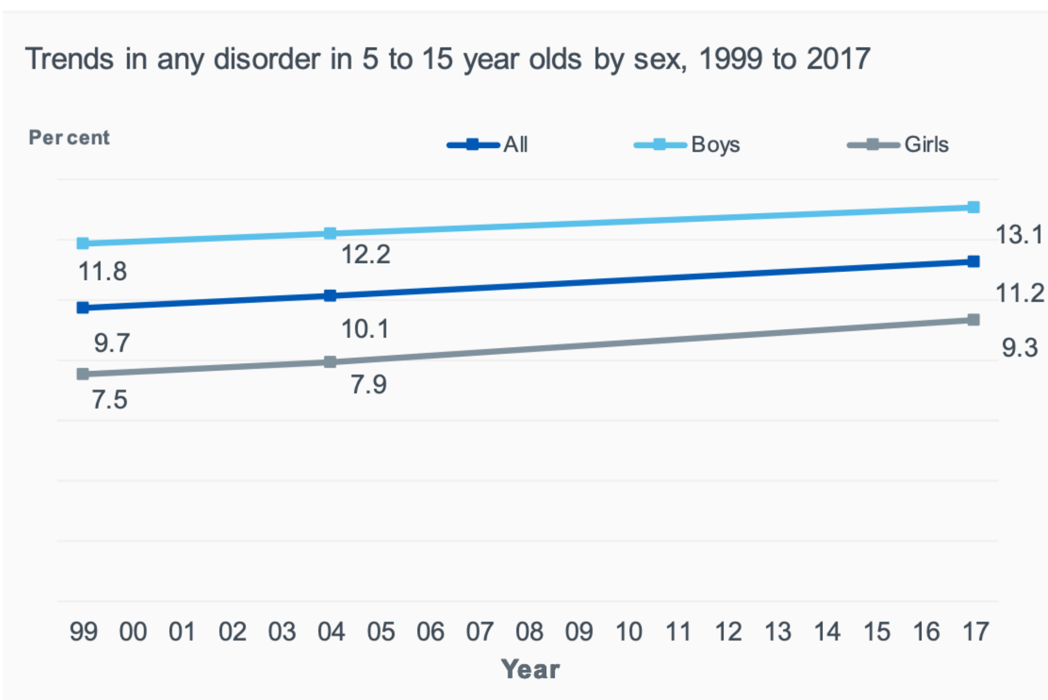

Figure 7: Trends in any disorder in 5 to 15 years old by sex, 1999 to 2017 (Sadler et al., 2018)

And more recently, during the Covid-19 pandemic, parents, and caregivers in the United Kingdom (n = 2673) of school-aged children and young people aged 4 to 16 years old took an online survey regarding their child's mental health at two periods between March and May 2020, during early lockdown (Waite et al., 2021). The study looked at changes in emotional symptoms, behavioral issues, and hyperactivity/inattention (Waite et al., 2021). The findings revealed a 10% rise in individuals experiencing probable emotional symptoms, a 20% increase in hyperactivity/inattention, and a 35% increase in behavior difficulties among preadolescent children (Waite et al., 2021). Adolescents, on the other hand, experienced modest alterations (4% and 8% increases in hyperactivity/inattention and conduct issues, respectively), with a minor drop in emotional symptoms (representing a 3% reduction) (Waite et al., 2021). Children and adolescents from low-income households, as well as those with special educational needs and/or neurodevelopmental disorders, had elevated symptoms at both time points (Waite et al., 2021). However, the lack of longitudinal data during a pandemic limits our understanding of how the national lockdown (which included school closures for most children) was linked to changes in mental health (Waite et al., 2021).

In China, 22.6% of 2330 young people reported in a survey elevated depressive symptoms and 18.9% reported elevated anxiety symptoms during the lockdown (Xie et al., 2020). These findings imply that major viral infections, like other traumatic experiences, may have an impact on children's mental health (Xie et al., 2020). One drawback is that our present investigation was unable to determine if these consequences will be sustained following the COVID-19 pandemic (Xie et al., 2020).

Providing definitive answers on longitudinal changes in mental health is extremely difficult from a scientific standpoint since changes in the diagnostic standards, variations in the techniques of evaluation, and adjustments to the official reporting procedures frequently have an impact on comparisons of the rates of disorder at various time points (McMartin et al., 2014). Only a few studies have used the same equipment at each time point (McMartin et al., 2014). Furthermore, measuring variations may contribute to disagreements in studies concerning changes in the prevalence of mental disease in the general population (McMartin et al., 2014). Certain studies are determined by a health professional's diagnosis of a condition, whilst other studies are determined by parent or teacher reports of symptoms or in-person interviews (McMartin et al., 2014). When mental health awareness increases, parents and other adults who care for children are more likely to notice mental health disorders and refer them to professionals (McMartin et al., 2014). This may explain studies' results of rising rates of child and adolescent mental illness, as well as recent increases in the number of mental health concerns diagnosed (McMartin et al., 2014). More than ever, it is important to keep studying and promoting mental health in youth as studies tend to disagree with one another.

2.4 Understanding of trends in adolescent mental health

To determine if initiatives to promote children's mental health are effective, it is crucial to keep monitoring mental health over time (Collishaw et al., 2004). Studying changes in child mental health is also crucial for planning future investments in mental health care and understanding whether outcomes for children with mental health disorders are improving or worsening (Collishaw et al., 2004).

Teenagers' lifestyles are a trend that has changed dramatically. Sleep, physical activity and sedentary behavior, obesity, screen time, and drug use are all potential contributory risk factors for teenage depression and anxiety (Collishaw et al., 2004). These factors have all changed over time. Between 2005 and 2015, young people are sleeping less and in worse quality (<8h 5.7% to 11.5%), are more likely to be overweight (obesity from 3.8% to 7.3%), and spend significantly more time engaged in screen-based activities such as social media interaction than earlier generations (Patalay & Gage, 2019). Simultaneously, recent decades have witnessed more beneficial developments in young people's habits, most notably a global decrease in youth smoking (9.2% to 2.9%), alcohol consumption (52.1% to 43.5%), cannabis (4.6% to 3.9%) and sexual activity (2% to 0.9%) (Patalay & Gage, 2019). More research is needed to determine how much these changes have led to (or reduced) changes in juvenile mental health (Patalay & Gage, 2019).

Bullying, loneliness, and a lack of supporting connections can all contribute to difficulties in social interactions (Collishaw et al., 2004). This can create mental health challenges and disorders in children's social functioning (Collishaw et al., 2004). Changes in the incidence of bullying and other social interaction challenges are important to investigate as potential contributors to youth mental health trends, especially with the widespread implementation of antibullying programs in schools in many countries (Chester et al., 2015). Evidence shows that bullying victimization has declined in many countries and that countrywide school-based antibullying programs may be connected to reductions in traditional bullying in areas where these programs have been implemented (Chester et al., 2015). On the other hand, the prevalence of cyberbullying has grown over time (Kessel Schneider et al., 2015). Between 2006 and 2012, cyberbullying increased from 15% to 21% (Kessel Schneider et al., 2015). Other issues with children's social interactions, such as loneliness and isolation, continue to be prevalent and have not changed significantly over a period of 24 years between 1989 and 2013 (Lempinen et al., 2018). Additionally, it has been proposed that rising academic standards, demands, and workloads have a negative impact on young people's mental health (Collishaw et al., 2004). However, there is conflicting information about trends in children's perceptions of academic pressure and how they relate to developments in mental health (Collishaw et al., 2004).

While family demographic trends (e.g., the long-term rise in parental separation) are commonly documented, less is known about family dynamics trends (e.g., parental discord, parenting, and parent-child relationship quality) (Collishaw et al., 2004). This is due to the lack of comparable family life metrics in most cross-cohort comparisons (Collishaw et al., 2004). There is no evidence to support the idea that parenting standards have generally declined, even though there is some indication that young people are now more concerned than ever about their families' stability and connections (Sweeting et al., 2010). For instance, when teenagers from 1986 and 2006 were compared, those from the later cohort reported having parents who were equally supportive of them emotionally and who spoke with them well and spent meaningful time with them (Sweeting et al., 2010). There is strong evidence that mental health issues like anxiety and depression run in families, and this risk is passed on socially as well as genetically (Tully et al., 2008). A study done on both nonadopted and adopted adolescents had a noticeably higher risk of severe depression and disruptive behavioral problems when either the mother or both of their parents had major depression (Tully et al., 2008). Except for ADHD in adopted adolescents, paternal sadness had no primary influence on any adolescent mental condition and did not statistically indicate an increased likelihood of disorders in either nonadopted or adopted adolescents (Tully et al., 2008). Increases in young people's emotional difficulties in one generation have the potential to raise the likelihood of emotional problems in the next generations, which is a significant implication (Tully et al., 2008).

It is important to analyze not just historical patterns in the course and prognosis of core mental health symptoms, but also the wider functional implications of mental health disorders on the health and development of teenagers (Collishaw et al., 2004). Following these trends can help us understand which school initiatives are working in the long run and how we can prevent mental health disorders at a young age.

2.5 Technologies that help support mental health

Technology allows for the delivery of mental health treatments remotely and on a large scale (Figueroa & Aguilera, 2020). Online mental health solutions need to be accessible, inexpensive, and suitable for a wide range of people with different ages, languages, and levels of digital proficiency (Figueroa & Aguilera, 2020). Scaling up the delivery of private mental health services to patients across a variety of platforms, from tele-mental health to mobile treatments like apps and text messaging, requires a digital mental health revolution (Figueroa & Aguilera, 2020).

Most of the research on teleconferencing services found that they are as helpful as in-person services for conditions such as depression, post-traumatic stress disorder, and anxiety disorders (Ralston et al., 2019). This strategy has had some success in China during the pandemic as they effectively offered online psychological therapy, and self-help was widely pushed out by mental health specialists in medical institutions, universities, and academic societies (Liu et al., 2020). Before the pandemic, barely one in ten patients in the United States used telemedicine, and 75% were ignorant of telehealth alternatives or how to access them (Power, 2019).

Applications for mental health have been beneficial in reducing anxiety and depression symptoms (Firth et al., 2017). Over 10 000 consumer-accessible meditation, health, and mindfulness applications are available in app stores, however, many of them are not supported by research

(Larsen et al., 2019). In a survey performed on the 73 top-rated applications from the Play Store and iTunes, 33% of the methods were not supported by literature searches (Larsen et al., 2019). None of the apps addressed certification or accreditation procedures, and just one app cited published literature (Larsen et al., 2019). It is critical that mental health professionals only endorse apps with solid supporting data. Additionally, even though many individuals download mental health applications, research reveals low rates of long-term usage (Baumel et al., 2019).

Text messaging is an appropriate medium for those with limited digital literacy and underprivileged groups (Silver, 2019). Health and government entities can automate the delivery of trustworthy information via text messages (Figueroa & Aguilera, 2020). Scaling up the distribution of information to patients and the public might help free up the already overworked, understaffed, and underfunded public health agencies (Figueroa & Aguilera, 2020).

Social media can be a great tool for information and misinformation at the same time. Research demonstrates that many people with mental illness are increasingly using social media to share their experiences and look for information and guidance (Naslund et al., 2019). On the other side, it can aggravate depression and anxiety symptoms due to negative social comparisons and the spread of upsetting information (Primack et al., 2017). Large social media platforms are purportedly employing artificial intelligence (AI) to delete incorrect content or conspiracy theories regarding the Covid-19 pandemic and broadcast trustworthy information, such as that established by the World Health Organization (Douek, 2020). According to a recent survey of over 8,000 people from six countries, one-third reported seeing a considerable quantity of inaccurate or misleading Covid-19 material on social media or messaging platforms (Nielsen et al., 2020).

Virtual agents, also known as chatbots, are new tools health professionals can use to give out the right information to the intended users. Eastern Health, the biggest integrated health authority in Newfoundland and Labrador, are utilizing conversational AI, a tool that gives Eastern Health workers, clinicians, and their families access to mental health help and services 24/7 (McCarthy, 2021). Generally, people searching for mental health assistance require a bit of assistance and this is where conversational AI comes into play (McCarthy, 2021). The employee virtual assistant (Eva) speaks to individuals clearly to direct them to the appropriate options, whether that's self-help, virtual services, or even rapid emergency treatment (McCarthy, 2021). EVA also alleviates some people's apprehension about disclosing their mental health difficulties (McCarthy, 2021). Some people feel better at ease discussing their problems with a non-human bot (McCarthy, 2021). Watson's strong privacy safeguards allow users to feel confident that their privacy will be always secured (McCarthy, 2021). This machine learning tool can be modified at any time by the administrators to provide different information as they see fit and constantly learns from users' new questions.

In sum, the proper infrastructure must be in place to deliver digital interventions safely, without violating personal privacy, and with the danger of data breaches minimized. (Moussa et al., 2017) Apps and texting must be not only useful, but also safe, secure, and responsible, in the same way that therapists are committed to principles of responsible practice and confidentiality (Moussa et al., 2017). Finally, technologies that support mental health should be always accessible, inexpensive, and suitable for a wide range of people of different ages, languages, and levels of digital proficiency (Figueroa & Aguilera, 2020).

3. Hypothesis Development

Now that we are collectively aware of the number of mental health issues teenagers face and their problems, and the intervention methods healthcare workers use to help these students face these issues, we are now going to pivot to the method we will be using in this paper. Curve clustering will be useful to detect temporal patterns and classify teenagers' emotions to help healthcare professionals help these students. This subsection sets the table for the method of our study by establishing how researchers have found a way to cluster data on a longitudinal basis and how we'll be using it in our study to represent the different clusters of students.

3.1 Introduction to curve clustering concepts

Clustering is an effective method for explanatory pattern recognition analysis, grouping, decision-making, and machine learning in applications such as data mining, document retrieval, picture segmentation, and pattern classification (Ramsay & Dalzell, 1991). Clustering is accomplished by categorizing objects in the same cluster as being more similar to one another than those in different clusters based on some predetermined criterion (Montero & Vilar, 2014). Curve clustering on the other hand operates on curves as a unit (Cheam & Fredette, 2020). To generate clusters given a set of unlabeled curves, curve clustering requires a clustering technique, and the clustering method chosen relies on both the kind of dataset and the specific goal and application (Cheam & Fredette, 2020).

Many fields are interested in cluster analysis which is more often than not performed on some time series data (Montero & Vilar, 2014). Some typical applications where time series similarity searching is justified are : finding stocks that act similarly, finding products with similar selling patterns, determining nations with comparable population growth, identifying regions with similar temperatures, etc. (Montero & Vilar, 2014). Time series cluster analysis emerges naturally in many domains such as economics, finance, health, ecology, environmental studies, engineering, and many others (Montero & Vilar, 2014). Our study is in the domain of mental health in high school students for nine months. As such, the grouping of time series is crucial to our subject.

Similarity and dissimilarity need to be understood to measure two objects. The idea of dissimilarity is particularly complicated in the context of time series data because they overlook the dependency link between values (Montero & Vilar, 2014). Curve clustering needs a defined clustering technique and there are several methods for comparing time series. Finding the appropriate dissimilarity measure is dependent on the nature of the clustering, i.e., defining what the objective of the grouping is (Montero & Vilar, 2014). In a paper "on the importance of similarity characteristics of curve clustering and its applications", the authors attempt to establish a new taxonomy of curve clustering by outlining four key resemblance traits that may be used to classify different techniques of comparing time series (Cheam & Fredette, 2020). Most time series clustering algorithms are based on generic processes (e.g., k-means or hierarchical clustering) that employ a set of dissimilarities built expressly for dealing with time series (Liao, 2005).

The R package TSclust was created by combining a variety of dissimilarity metrics to accomplish time series clustering (Montero & Vilar, 2014). The package includes frequently used dissimilarity measures such as complexity-based approaches, model-based approaches, feature-based approaches, and prediction-based approaches (Montero & Vilar, 2014). There are 27 different methods and counting in this package and each of them has its own similarities and differences.

3.2 Shortcomings

The main issue with curve clustering is selecting the appropriate metric to determine the dissimilarity between two-time series data (Montero & Vilar, 2014). Time-series data tend to be very big, therefore they've got dimensionality working against them (Kalpakis et al., 2001). On the same wavelength, in classification literature, several distinct metrics of pairwise similarity and dissimilarity have been presented (Kalpakis et al., 2001). TSclust is a package that attempts to combine many time series dissimilarity criteria into a single software package to test and evaluate their behaviors in clustering (Montero & Vilar, 2014). Except for the PDC package (Brandmaier, 2015), which is mostly focused on permutation distribution clustering, there are no prior packages available that tackle the challenge of clustering time series (Montero & Vilar, 2014). For this purpose, TSclust is constantly being updated to introduce additional dissimilarity criteria and clustering utilities into the time series framework (Montero & Vilar, 2014). Nonetheless, the growing number of references and applications in other disciplines support the demand for a bundle containing these capabilities (Montero & Vilar, 2014).

3.3 Types of approaches in TSclust

There are four different types of approaches described by Montero and Jose who built this TSclust package: model-free approaches, model-based approaches, complexity-based approaches, and prediction-based approaches.

The model-free approach is measured by the distance between X_T and Y_T, meaning the distance between two points, by traditional metrics such as Minkowski distance, Fréchet distance, Dynamic time warping distance, Correlation-based distances (in our case), autocorrelation-based distances, etc... (Montero & Vilar, 2014).

The model-based dissimilarity implies that the underlying models are produced by X_T and Y_T following invertible ARIMA (Autoregressive integrated moving average) models (Montero & Vilar, 2014). This model is applied to time series data to either better comprehend the data or forecast future points in the series (Kalpakis et al., 2001). The objective in this situation is to fit an ARIMA model to each dataset and then measure the dissimilarity between the fitted models (Montero & Vilar, 2014). The structure is normally estimated using AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), or least squares estimators (Montero & Vilar, 2014). The three models in this package are Piccolo distance, Maharaj distance, and Cepstral-based distance (Montero & Vilar, 2014).

The complexity-based approach measures the similarity of two-time series by using the degree of shared information by both time series rather than specific serial properties or knowledge of underlying models (Montero & Vilar, 2014). This measure needs to be estimated using the Kolmogorov complexity notion (Montero & Vilar, 2014). The three methodologies are compression-based dissimilarity measures, permutation distribution clustering, and a complexity-invariant dissimilarity measure (Montero & Vilar, 2014).

Finally, the prediction-based approach focuses on future forecasts which means that two time series are comparable if their projections of the future are similar (Montero & Vilar, 2014). An instance in which the true interest of clustering is directly dependent on the features of the forecasts would be the sustainable development challenge (Montero & Vilar, 2014). This method is still new to the package. Different approaches produce different results. Many of these models can be interchanged depending on the goal of the study.

3.4 Correlation-based Dissimilarity

In this study, the goal is to find clusters of similarities in the emotions of high school students over time. In the analysis, the model-free approach from the R package TSclust is used, specifically the diss.COR is also known as *Correlation-based (Montero & Vilar, 2014)*. This measurement "computes dissimilarities based on the estimated Pearson's correlation of two given time series" (Montero & Vilar, 2014). Correlation-based dissimilarity was used because we are interested in the time-lock pattern. We want to observe patterns visible when data points are taken at the same time.

Pablo Montero Manso and José Antonio Vilar (Montero & Vilar, 2014) explain the following:

> A first and simple dissimilarity criterion is to consider the Pearson's correlation factor between X_T and Y_T given by:
>
> $$COR\left(\boldsymbol{X}_T, \boldsymbol{Y}_T\right) = \frac{\sum_{t=1}^{T}\left(X_t - \overline{\boldsymbol{X}}_T\right)\left(Y_t - \overline{\boldsymbol{Y}}_T\right)}{\sqrt{\sum_{t=1}^{T}\left(X_T - \overline{\boldsymbol{X}}_T\right)^2}\sqrt{\sum_{t=1}^{T}\left(Y_t - \overline{\boldsymbol{Y}}_T\right)^2}},$$
>
> with $\bar{X}_T$ and $\bar{Y}_T$ the average values of the serial realizations X_T and Y_T respectively. Golay, Kollias, Stoll, Meier, Valavanis, and Boesiger (Golay et al., 1998) construct a fuzzy k-means algorithm using the following two cross-correlation-based distances:
>
> $$d_{COR.1}\left(\boldsymbol{X}_T, \boldsymbol{Y}_T\right) = \sqrt{2\left(1 - COR\left(\boldsymbol{X}_T, \boldsymbol{Y}_T\right)\right)}, \quad \text{and} \quad d_{COR.2}\left(\boldsymbol{X}_T, \boldsymbol{Y}_T\right) = \sqrt{\left(\frac{1 - COR\left(\boldsymbol{X}_T, \boldsymbol{Y}_T\right)}{1 + COR\left(\boldsymbol{X}_T, \boldsymbol{Y}_T\right)}\right)^{\beta}}, \text{ with } \beta \geq 0.$$
>
> Note that $d_{COR.2}$ becomes infinite when COR $(X_T, Y_T) = -1$ and the parameter β allows regulation of the fast decreasing of the distance.

3.5 Hypothesis

The focus of analysis in this research paper is curve clustering. Curve clustering is fundamentally finding similar characteristics in data based on their curves and finding meaningful groups where we can maximize their similarities and dissimilarities (Cheam & Fredette, 2020). For a collection of curves, two curves are said to be comparable if they have the same curvature, whether scale-invariant or not (Cheam & Fredette, 2020). Another approach may be based on distance rather than semblance, in which two curves are regarded as similar if their values are near at each time point (Cheam & Fredette, 2020). In our case, we will be observing the curves through time and measuring their clusters with the correlation-based dissimilarity method. This method is interesting as it follows the surveys in a time-lock pattern, meaning it assumes students are answering the survey on the same day every time. This is a limitation as the method works perfectly in our case except that it wasn't evaluated on the same day. We have accounted for this in the methodology, as when cleaning the data, depending on whenever the students first answered their surveys, it will be considered as the first week and so on. With curve clustering, we will be able to find clusters of similarities and/or differences between students and bring suggestions on how to increase students' well-being. A reminder of our research question: is it possible to find clusters of students over time using curve clustering based on their longitudinal similarities/differences in their emotions and find characteristics that differentiate these groups?

Based on our research questions, our hypothesis:

H0: With curve clustering analysis techniques over time, we anticipate finding different meaningful clusters of students.

We will analyze to find that the level of energy and motivation are the main characteristics explaining these groups.

Chapter 3: Methodology

Study design

In collaboration with the not-for-profit organization *Dis-Moi*, we collected a longitudinal survey on the mental health of high school students on either a voluntary or in-class basis from August 2020 to May 2021 for 20 weeks. The goal of these surveys is to help school educators understand and give each student the proper care they need based on how they perceive their situation and the emotions they are feeling. Finding clusters can help differentiate not only by gender and school year but by lifestyle and emotional level these students are facing. This can help school educators understand their students and give them proper guidance and tools adapted to the survey and the cluster they fall into.

This research was approved by the HEC Montreal Ethics Committee (CER: Comité d'éthique de la recherche) whose mandate is to ensure that participants in the study conducted by the researcher are both voluntary and consenting and that they have all the information they need to make their decision.

Research participants

All participants in this study were high school students from ages 11 to 16. Seven Quebec schools were involved in this study, four of which had to answer the mental health survey in class and three of which answered it voluntarily.

The study was conducted on 1226 students with diverse backgrounds who were given a survey to complete via an online platform and the participants either answered it within their own free time or during class time. After manipulating and cleaning the database, only 85 students' IDs were finally used for the study. This was due to three main factors. First was the low answer rate as we calculated a minimum of 30% answer rate, which would be answering the survey 7 times throughout the school year, reducing the sample size down to 241 students. The second was the regularity in answering the surveys as some students answered multiple surveys on the same day making their following answers irrelevant as we are measuring their emotions through time, and the third reason is due to low variance as the main question we are evaluating (emotion) was always the same every week. We will evaluate this group later in the study.

Measures

The surveys were conducted to measure the participants' experiences and their level of emotion each week. The questionnaires' questions were structured to understand different factors in the students' lives and how they might affect their emotional state that week. The students had to self-evaluate their motivation and ability for 9 different questions (not given all at once every week except for the first question).

These students had to answer a series of mental health questions which differed every week, except for one weekly question which was "Today, I feel..." followed by a list of 7 emotions to choose from (Angry/Irritated 🤬, Depressed 😥, Sad/Depressed 😟, Anxious 😰, Top! 👌, Good 😃, Okay 😐).

These are the questions that have been collected:

Table 1: Variables

High school level	Secondary 1 Secondary 2 Secondary 3 Secondary 4 Secondary 5
Gender	Female Male
Q1: Level of emotion	1: Depressed 😥 2: Sad, down 😟 3: Anxious 😰 4: Angry, irritated 🤬 5: Fair 😐 6: Good 😃 7: Great! 👌
Q2: Level of physical activity	0: Never 1: 2 times 2: 3-5 times 3: More than 5 times
Q3: Average level of sleep each night	0: 3-4 hours 1: 5-6 hours 2: 7-9 hours 3: 10 hours or more
Q4: Average level of energy	0: Low 10: High
Q5: Level of stress/anxiety	0: Very low 10: Very high
Q6: Level of relationship with friend	0: Terrible 10: Great
Q7: Level of motivation at school	0: Very low 10: Very high
Q8: Time in front of a screen excluding schoolwork	0: Less than 1 hour per day 1: 1-2 hours per day 2: 3-4 hours per day 3: 5 hours or more per day
Q9: Level of communication with family	0: Not at all true 10: Absolutely true

Analysis

To test our hypothesis, a python script and an R script were created to clean, curve cluster, analyze, and conclude.

The data was first transformed from categorical values to numerical values for simplicity of handling. We can justify the first question as being on an ordinal scale as it establishes an orderly relationship between the emotions perceived by students. Without being able to quantitatively assess the distance between each emotion, we can use this scale to understand their level of distress or happiness. This was based on Russell's Circumplex Model and how we perceive the deactivation quadrant as worse than the unpleasant quadrant (Russell, 1980). In our case, we were able to order the emotions of the students from the survey in this order: Depressed > Sad, down > Anxious > Angry, irritated > Fair > Good > Great. As these answers were pre-established by the organisation, they do not correspond perfectly to the model. This was done to help with further analysis and will be discussed in the limitations.

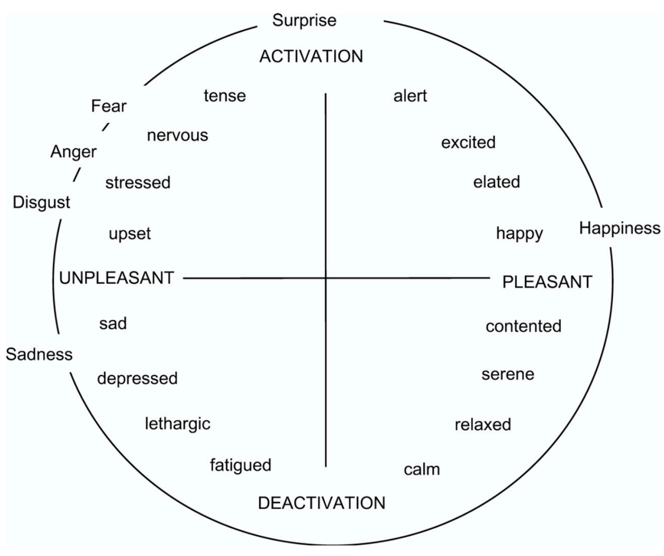

Figure 8: Russell's (1980) Circumplex Model

The answer rate was then evaluated. Respondents with an answer rate of 30% and lower were excluded from the study as there were not enough values to correctly identify their cluster. This reduced the sample to 241 students. With the remaining respondants, all missing values were imputed by most frequent meaning by mode. Mode imputation is the process of replacing all instances of missing values (NA) inside a table with the mode, which is the most common value or category. Since our dataset is categorical and isn't a highly skewed class distribution, this would be the best method.

Afterward, the regularity of answers was examined. The students had to answer the survey weekly but also had access to all the surveys they didn't answer in the past. This could allow them to answer the current week but also the ones of the past weeks on the same day. This reduces credibility as their answers weren't factual and the emotion perceived for the past weeks wouldn't be accurate. To resolve this issue, we evaluated when they answered their surveys based on the timestamp of completion. The extra surveys that they answered on the same day were deleted, keeping that week's survey as the most up-to-date version.

The variance was then investigated as this determined how different their answers were from week to week. It was established that if there was a variance of 25% or less, meaning that 75% of the time and more the student answered the same emotion on their survey, meaning that they might not be taking the study seriously, or that they felt quite stable throughout the study. We will later evaluate this group as a cluster in itself.

The goal is to visualize and demonstrate if there are similarities or differences between the clusters. A cluster analysis and a multinominal regression were completed to understand which variables significantly impacted the clusters. We were able to identify personas and find the characteristics of each cluster.

Chapter 4: Results

The methodology enabled us to collect a great amount of explicit data. Descriptive statistics and statistical analysis following curve clustering methodology were done to extract three different behavioral patterns. To do our clustering, we used only the first question (Q1: level of emotion), and then we were able to outline these clusters with the other characteristics collected in the study.

Descriptive Statistics

The following table 2 is a descriptive analysis summary of our dataset before any cleaning and implementation of clusters were made. To start off, the count of every value is different except the gender and the high school level. That is because there are 1226 students in this dataset, but the survey was conducted in a span of 21 weeks meaning the questions Q1 to Q9 were answered multiple times. Q1 was answered every week and all the other questions were answered every two weeks making the response rate lowered. As we can tell from the maximum, the range of answers differed for each question as there were scales from 0 to 2, 1 to 5, 1 to 7, 0 to 3, and 1 to 10. We have standardized our analysis to account for this. Gender has 3 levels since it is female, male, and other. Further in the study, there won't be any "other" in our final clustering.

Table 2: Descriptive Statistics

	Gender	High School Level	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Mean	0.67	2.85	5.23	1.72	1.77	2.58	4.85	6.86	5.18	1.73	5.26
Standard Deviation	0.55	1.35	1.44	0.92	0.67	2.58	2.98	2.45	2.83	0.85	3.09
Median	1	3	5	2	2	6	5	7	5	2	5
Mode	1	4	5	2	2	7	7	5	5	2	5
Kurtosis	-0.73	-1.23	1.41	-0.90	0.88	-0.71	-1.03	-0.15	-0.84	-0.62	-1.07
Skewness	0.05	0.05	-1.30	-0.11	-0.73	-0.15	-0.06	-0.61	-0.06	-0.17	-0.11
Range	2	4	6	3	3	10	10	10	10	3	10
Minimum	0	1	1	0	0	0	0	0	0	0	0
Maximum	2	5	7	3	3	10	10	10	10	3	10
Count	1226	1226	3796	1657	1664	1659	2109	2165	2154	2147	1667

Our dataset was cleaned based on our methodology. By using python, we took away all respondents with low answer rates as we calculated a minimum of 30% answer rate, which would be answering the survey 7 times throughout the school year. Then, we evaluated the regularity in the answers to the surveys as some students answered multiple surveys on the same day making their following answers irrelevant as we are measuring their emotions through time. Afterward, we evaluated low variance for the first question and kept it as its own cluster further in this study.

Clustering

After extracting and processing all the data, we performed a Cluster Analysis. Cluster Analysis is a statistical method used to divide the observations and data into different groups (Montero & Vilar, 2014). Many classification systems and algorithms for clustering analysis exists, but the k-means procedure seems to be the most relevant and robust hierarchical method (Montero & Vilar, 2014). The measurements between curves can be found in section 3.4 of this paper. To confirm our hypothesis, we based our cluster analysis on the participant's emotions following each week. (i.e., the level of emotion). Following the curve clustering definition, we wish to demonstrate that behaviors are similar within the same segment and behaviors are different between the segments.

To induce similarities within the pool of participants, we draw for every participant a curve of his emotions in a graph (Q1) where the x-axis represents the weeks and the y-axis, the level of emotion. We assume here that the distance between two emotions that follow each other is the same. We then group the participants who had similar patterns according to the emotions portrayed every week. The TSclust diss.COR R package was run to create clusters with the remaining 85 user IDs. We ran the R code with a varying number of clusters. We used only the first question of the study (level of emotion) to cluster since it was the only variable measured weekly. The input to the R code was the user ID's level of emotion based on the week. The output of the R code was the user IDs and their respective cluster number assigned. With these different results, we were able to establish that N=3, meaning three clusters, had visually the best curves. When observing 2 clusters (figure 9), there wasn't enough information to be collected and cluster 4 (figure 11) had too many crossovers in curves. This left cluster 3 (figure 10) the best option going forward.

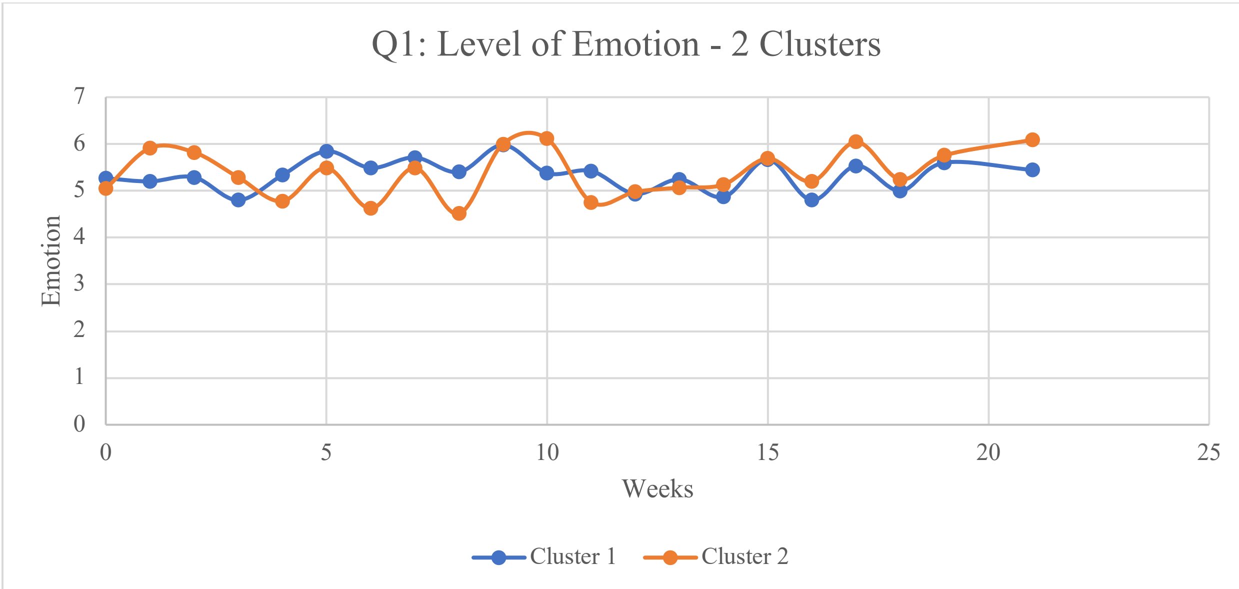

Figure 9: 2 clusters for Q1 (level of emotion) with results over 25% variance

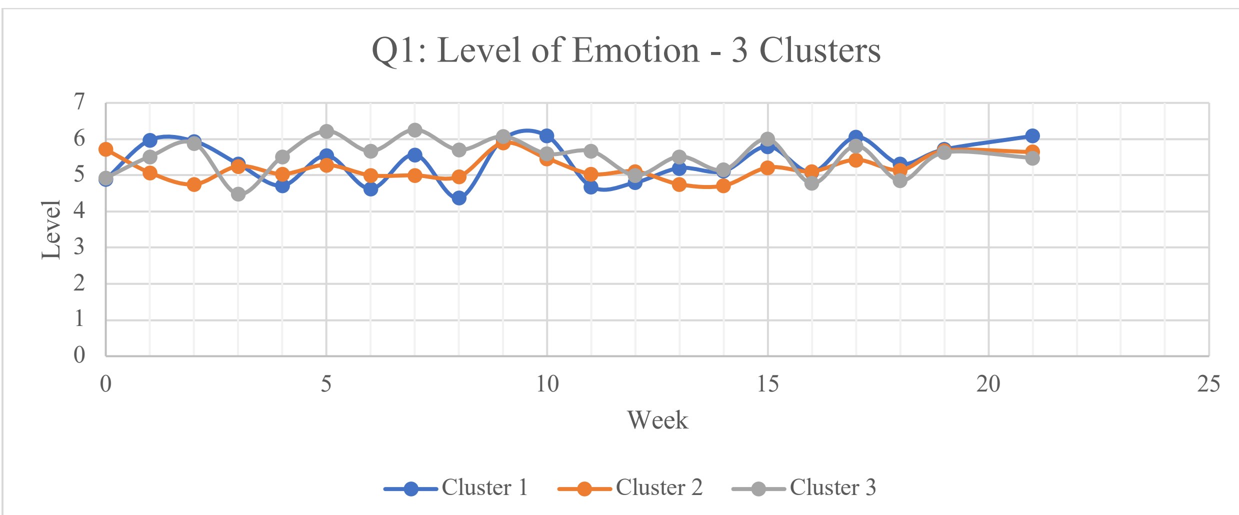

Figure 10: 3 clusters for Q1 (level of emotion) with results over 25% variance

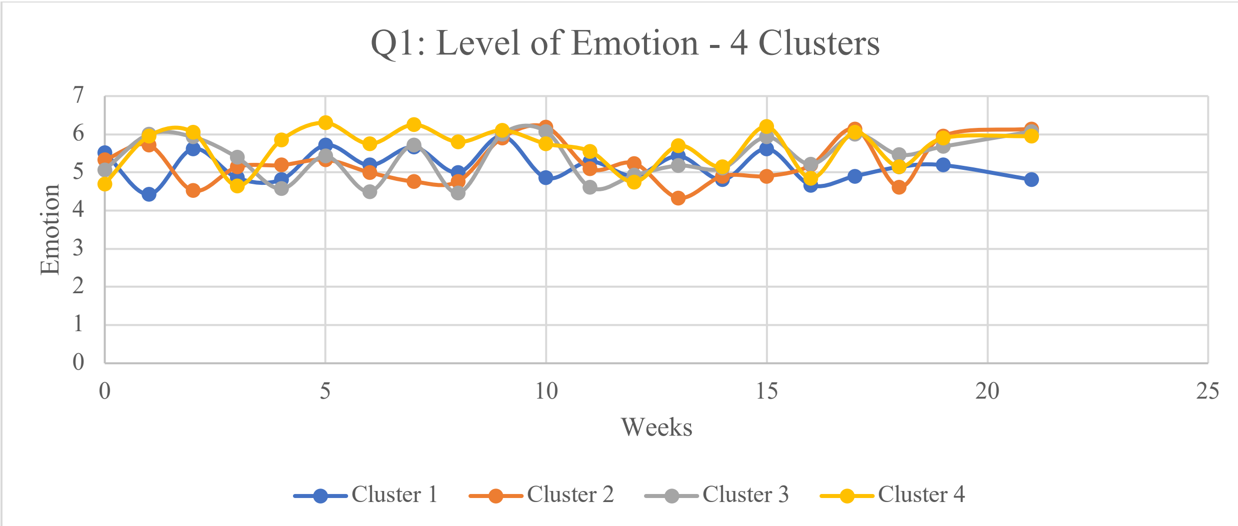

Figure 11: 4 clusters for Q1 (level of emotion) with results over 25% variance

Three clusters were created to describe the user's emotions. The composition was 33 students in cluster 1, 27 students in cluster 2, and 25 students in cluster 3. The user IDs were then mapped to their corresponding demographic, high school level, level of physical activity, average level of sleep each night, average level of energy, level of stress/anxiety, level of relationship with friends, level of motivation at school, time in front of a screen excluding schoolwork, and level of communication with family. Doing this gave us the possibility to create a persona for each cluster. The clusters were created based on student's level of emotions and the answers to all the other questions collected describes these clusters in more depth.

A fourth cluster (figure 12), composed of 12 students, was added to this comparison as the group with low variance. With these clusters, we can safely answer our hypothesis that with curve clustering analysis techniques over time, we find different meaningful clusters of students. Further, in the results, we end up removing this fourth cluster after our findings revealed no significant differences between the fourth cluster to the others.

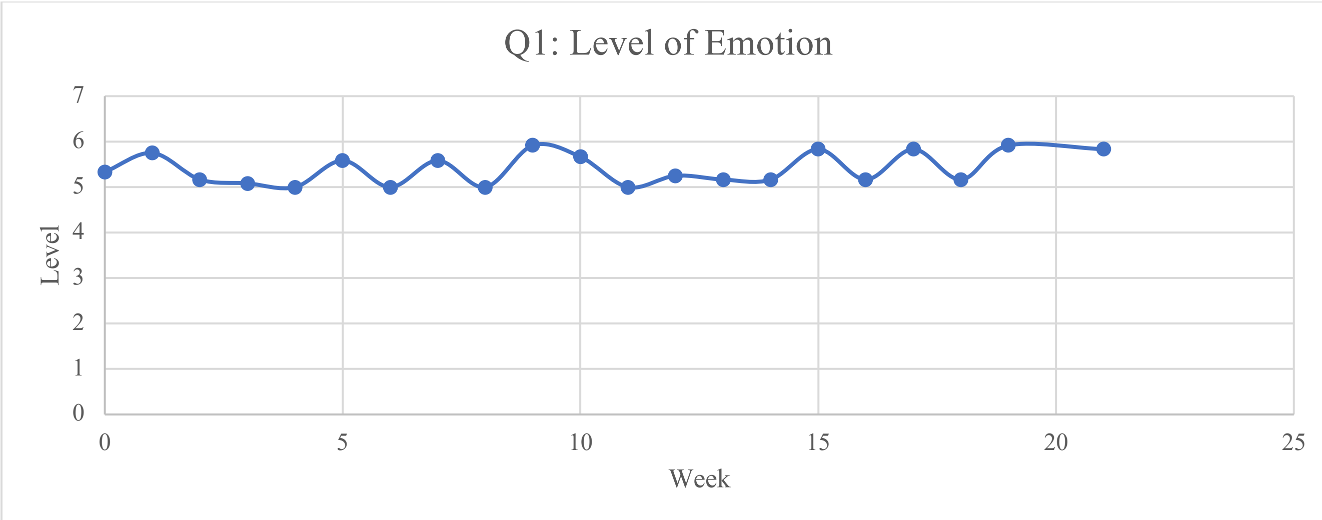

Figure 12: Cluster for Q1 (level of emotion) with a variance of 25% and less

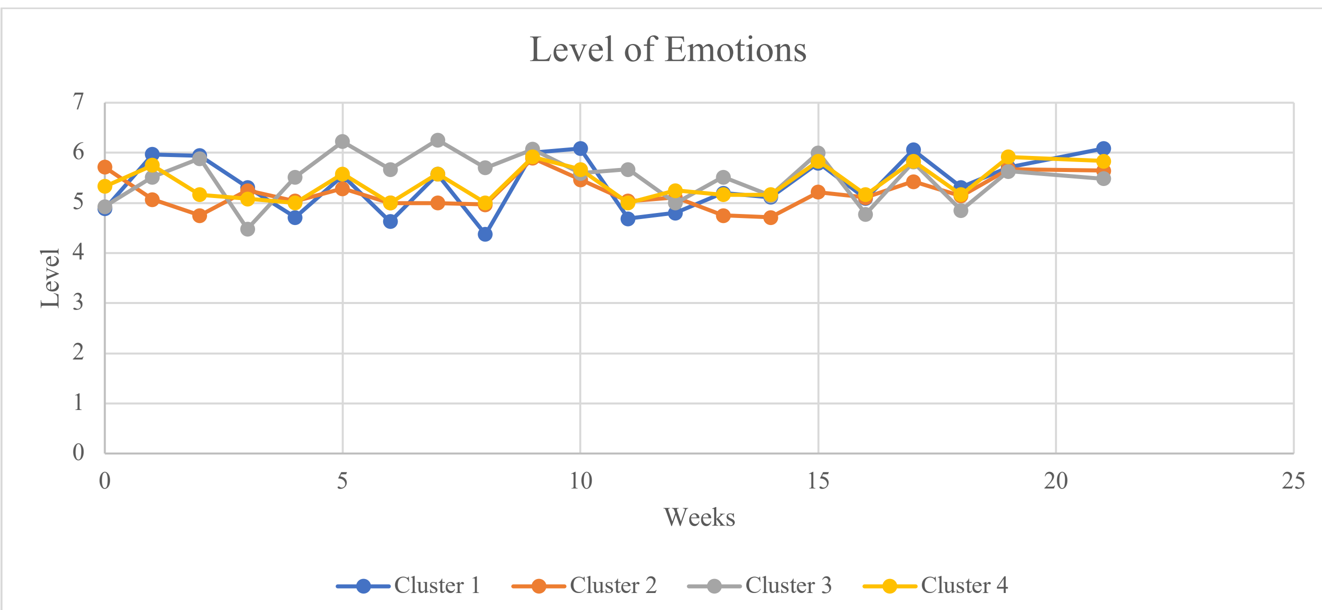

Figure 13: Four clusters combined

Figure 13 represents the four patterns for each cluster of students. It is noticeable that the lines are quite close together. This demonstrates that the clusters are quite similar, with only a few defining differences between them. To test statistically the differences between the clusters, we used a global test, a multinomial regression, a mean analysis, and a correlation analysis.

Defining the clusters

Once the clustering was finalized and the hypothesis answered, statistical tests were performed to understand which variables are the most significant in our model. This would allow us to find commonalities and/or differences between groups of students.

First, a cluster means analysis is performed to understand the similarities between each cluster. The first cluster is the biggest cluster with a count of 33 students, the second 27 students, the third 25 students, and the fourth cluster with 12 students.

Table 6: Cluster Means

Variable	**Cluster 1**		**Cluster 2**		**Cluster 3**		**Cluster 4**	
	Mean	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
High School Level	2.27	0.98	2.3	1.07	2.2	1.12	2.55	1.04
Gender	0.79	0.42	0.89	0.32	0.52	0.51	0.64	0.5
q1	5.38	0.38	5.18	0.48	5.54	0.34	5.43	0.15
q2	1.66	0.73	1.64	0.72	2.01	0.71	2.02	0.81
q3	1.87	0.48	1.65	0.52	2.01	0.55	2.02	0.46
q4	5.24	2.35	5.18	1.65	6.93	1.98	5.86	1.77
q5	4.93	2.05	5.79	2.25	3.98	2.33	4.11	2.19
q6	6.86	1.48	6.59	1.93	7.6	1.76	6.87	1.77
q7	5.56	1.95	4.52	1.97	6.19	2.13	5.35	1.79
q8	1.63	0.67	1.96	0.58	1.61	0.7	1.52	0.67
q9	5.5	2.62	4.82	2.33	6.35	2.65	6	2.16

A simple mean analysis is done to find the average value of each variable depending on the cluster they belong to, and its standard deviation as shown in table 6. With this information, we get a clearer view of the impact of each value between the clusters. As seen above, cluster 1 is very average. It does not have any of the highest or the lowest values in all the clusters.

Cluster 2 has the most women in all the clusters and the highest q5 and q8, meaning high stress/anxiety and high screen time after schoolwork. They also had the lowest q1 (level of emotion), q2 (amount of physical activity), q3 (amount of sleep), q4 (level of energy), q6 (relationship with friends), q7 (motivation at school), and q9 (communication with family). This means they had the lowest level of emotion, did the least physical activity, slept the least, had the least amount of energy, had worse relationships with friends, had the lowest level of motivation in school, and had worst communication with family than the other clusters.

Cluster 3 has the highest q1 (level of emotion), q3 (amount of sleep), q4 (level of energy), q6 (relationship with friends), q7 (motivation at school), and q9 (communication with family). This means they had the highest level of emotion, slept the most, had the highest amount of energy, had better relationships with friends, had the highest motivation in school, and had good communication with family. This cluster also scored the lowest in q5 and q8, meaning low stress/anxiety and low screen time after schoolwork.

Cluster 4 represents the 12 students with low variance. This cluster had the average highest high school level, the highest q2 (amount of physical activity), and the lowest q8 (screen time after schoolwork). This meant that this cluster did the most physical activity and had the least amount of screen time.

Afterward, a correlation matrix, in table 7, helps us understand a few of the similarities between the variables. A correlation matrix is a table that displays the correlation coefficients for numerous variables in a matrix (Team, 2022). This demonstrates the link between all possible value pairs in a table and can be a powerful tool for identifying and visualizing data patterns (Team, 2022).

Table 3: Correlation Matrix

	q1	q2	q3	q4	q5	q6	q7	q8	q9
q1	1	0.14844	0.46189	0.66528	-0.60704	0.53154	0.66727	-0.29684	0.60042
		0.1489	<.0001	<.0001	<.0001	<.0001	<.0001	0.0033	<.0001
q2	0.14844	1	0.17251	0.3508	-0.26014	0.02217	0.17914	-0.16205	0.16719
	0.1489		0.0928	0.0005	0.0105	0.8302	0.0808	0.1147	0.1035
q3	0.46189	0.17251	1	0.55244	-0.36002	0.14306	0.55066	-0.17738	0.51966
	<.0001	0.0928		<.0001	0.0003	0.1644	<.0001	0.0838	<.0001
q4	0.66528	0.3508	0.55244	1	-0.55823	0.45403	0.73887	-0.25088	0.70828
	<.0001	0.0005	<.0001		<.0001	<.0001	<.0001	0.0137	<.0001
q5	-0.60704	-0.26014	-0.36002	-0.55823	1	-0.48242	-0.63302	0.2346	-0.51191
	<.0001	0.0105	0.0003	<.0001		<.0001	<.0001	0.0214	<.0001
q6	0.53154	0.02217	0.14306	0.45403	-0.48242	1	0.56998	-0.20025	0.4947
	<.0001	0.8302	0.1644	<.0001	<.0001		<.0001	0.0504	<.0001
q7	0.66727	0.17914	0.55066	0.73887	-0.63302	0.56998	1	-0.31062	0.66135
	<.0001	0.0808	<.0001	<.0001	<.0001	<.0001		0.0021	<.0001
q8	-0.29684	-0.16205	-0.17738	-0.25088	0.2346	-0.20025	-0.31062	1	-0.35045
	0.0033	0.1147	0.0838	0.0137	0.0214	0.0504	0.0021		0.0005
q9	0.60042	0.16719	0.51966	0.70828	-0.51191	0.4947	0.66135	-0.35045	1
	<.0001	0.1035	<.0001	<.0001	<.0001	<.0001	<.0001	0.0005	

In table 3, Pearson Correlation Coefficients, N = 96 and Prob > |r| under H0: Rho=0. We can observe that energy (Q4) is strongly correlated to motivation (Q7) and level of communication with family. Motivation (Q7) is negatively correlated with stress(Q5). This is interesting as cluster C, which scored the highest motivation, also scored the lowest stress levels.

Once that was completed, a global test was performed to understand the significance of each variable. To clarify, we treated cluster id as a three-modality dummy variable and performed a generalized logistic regression for each explanatory variable. Table 4 represents the global test for each variable individually to identify if they have a significant marginal effect on ClusterID where H0 means the variables have no effect on the ClusterID.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.1942	12	0.5989
Score	9.4528	12	0.6638
Wald	7.4588	12	0.8259

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
HighSchoolLevel	12	7.4588	0.8259

Analysis of Maximum Likelihood Estimates							
Parameter		ClusterID	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		A	1	4.0931	550.5	0.0001	0.9941
Intercept		B	1	1.0950	653.0	0.0000	0.9987
Intercept		C	1	3.9188	550.5	0.0001	0.9943
HighSchoolLevel	1	A	1	-2.3013	550.5	0.0000	0.9967
HighSchoolLevel	1	B	1	0.8509	653.0	0.0000	0.9990
HighSchoolLevel	1	C	1	-1.8393	550.5	0.0000	0.9973
HighSchoolLevel	2	A	1	-3.0516	550.5	0.0000	0.9956
HighSchoolLevel	2	B	1	-0.5842	653.0	0.0000	0.9993
HighSchoolLevel	2	C	1	-3.6311	550.5	0.0000	0.9947
HighSchoolLevel	3	A	1	-2.3013	550.5	0.0000	0.9967
HighSchoolLevel	3	B	1	0.5144	653.0	0.0000	0.9994
HighSchoolLevel	3	C	1	-2.1270	550.5	0.0000	0.9969
HighSchoolLevel	4	A	1	-4.0931	550.5	0.0001	0.9941
HighSchoolLevel	4	B	1	-0.5842	653.0	0.0000	0.9993
HighSchoolLevel	4	C	1	-4.3242	550.5	0.0001	0.9937

Figure 14: Individual effects related to the logistic regression to measure the effect of HighSchool

Table 4: Summary of global tests for each IV individually

DV	Effect	DF	WaldChiSq	ProbChiSq	
ClusterID	HighSchoolLevel	12	7.4588	0.8259	High School Level
ClusterID	Gender	3	9.118	0.0278	Gender
ClusterID	q1	3	9.5727	0.0226	Level of emotion
ClusterID	q2	3	5.2174	0.1566	Level of physical activity
ClusterID	q3	3	6.9773	0.0726	Average level of sleep each night
ClusterID	q4	3	10.599	0.0141	Average level of energy
ClusterID	q5	3	8.698	0.0336	Level of stress/anxiety
ClusterID	q6	3	4.58	0.2053	Level of relationship with friends
ClusterID	q7	3	8.265	0.0408	Level of motivation at school
ClusterID	q8	3	5.5884	0.1334	Time in front of a screen excluding schoolwork
ClusterID	q9	3	4.9722	0.1738	Level of communication with family

The reason behind the "high school level" having a DF of 12 is because it is a categorical variable with 5 levels (one of them being the reference) so 4 levels and 3 different intercepts for the 3 clusters of cluster ID. The rest of the variables have a DF of 3 as they are continuous or binary and have 3 intercepts for the 3 clusterIDs. We can conclude that *Gender*, *q1* (level of emotion), *q3* (average Level of sleep each night), *q4* (average level of energy), *q5* (level of stress/anxiety), and *q7* (level of motivation at school) seem to differ by cluster. With these results, we need to understand which variables have the most significant effect by doing a multinomial regression.

Multinomial logistic regression is a method for categorical data analysis (El-Habil, 2012). The logistic regression model may be adapted to include several explanatory variables (El-Habil, 2012). The independent variables can be either binary or continuous (El-Habil, 2012). The goal of this regression is to have a global test for the individually significant variables jointly which means identifying the most discriminant variables having a significant effect on Cluster ID.

Table 5: Global test for the individually significant variables jointly

Effect	**DF**	**WaldChiSq**	**Pr > ChiSq**	
Gender	3	3.3015	0.3474	
q3	3	2.4026	0.4932	
q4	3	8.193	0.0422	Average level of energy
q5	3	2.3986	0.4939	
q7	3	5.6046	0.1325	Level of motivation at school

Table 6: Pairwise comparison of effect of Q4 (energy) and Q7 (motivation) on ClusterID

DV	ClusterID	Ref_ClusterID	Effect	Estimate	StdErr	DF	tValue	Probt
ClusterID	B	A	q4	-0.01539	0.1337	90	-0.12	0.9086
ClusterID	C	A	q4	0.416	0.1429	90	2.91	0.0045
ClusterID	D	A	q4	0.1574	0.1753	90	0.9	0.3719
ClusterID	B	A	q7	-0.2811	0.141	90	-1.99	0.0493
ClusterID	C	A	q7	0.1609	0.1358	90	1.19	0.2391
ClusterID	D	A	q7	-0.0562	0.1775	90	-0.32	0.7522
ClusterID	C	B	q4	0.4314	0.15	90	2.88	0.005
ClusterID	D	B	q4	0.1727	0.1809	90	0.96	0.3421
ClusterID	C	B	q7	0.442	0.1555	90	2.84	0.0055
ClusterID	D	B	q7	0.2249	0.1878	90	1.2	0.2342
ClusterID	D	C	q4	-0.2587	0.1819	90	-1.42	0.1585
ClusterID	D	C	q7	-0.2171	0.1862	90	-1.17	0.2466

Table 5 and 6 represents the multinomial regression done on the significant variables of the global test. Comparing Q1 between clusters is not interesting since the clusters are formed based on Q1 itself. We can see in table 5 that q4 and q7 are the most significant variables and thus, we use them in our pairwise comparison to find which cluster has the highest impact. This multinomial regression gives us a better view of the similarity and differences between each cluster. As we can observe in table 6, there are no significant p-values for cluster D. This means that it is not distinguishable from the others, therefore cluster D is not a persona. Cluster 4 is those with low variance which would suggest that low variance is not related to energy or motivation. We can investigate further using a mean analysis to see the differences in values between each cluster.

In table 7, we have a summary of the mean analysis and the pairwise comparison. We can find differences between each cluster.

Table 7: Differences between each cluster

A vs B	A has higher motivation (Q7) than B (5.56 vs 4.52, p=0.0493)
A vs C	A has lower energy (Q4) than B (5.24 vs 6.93, p=0.00045)
B vs C	B has lower energy(Q4) and motivation(Q7) than C

Table 8: Personas

Cluster	**Cluster Name**	**Energy**	**Motivation**	**Description (based on average, not trend)**
A	Cluster 1	5.24	5.56	Somehow motivated
B	Cluster 2	5.18	4.52	Low energy, low motivation
C	Cluster 3	6.93	6.19	High energy, high motivation
~~D~~	~~Cluster 4~~	~~5.86~~	~~5.35~~	~~Average~~

With the results in table 8, we can now describe each segment of the cluster. As we understood earlier, we can conclude that cluster 4 has no significant differences with the other clusters.

In figure 15, we can describe cluster 1 as the average student. If we look back at table 9 with the mean analysis, between clusters 1, 2, and 3, cluster 1 never has the highest nor the lowest value. It is the most average cluster.

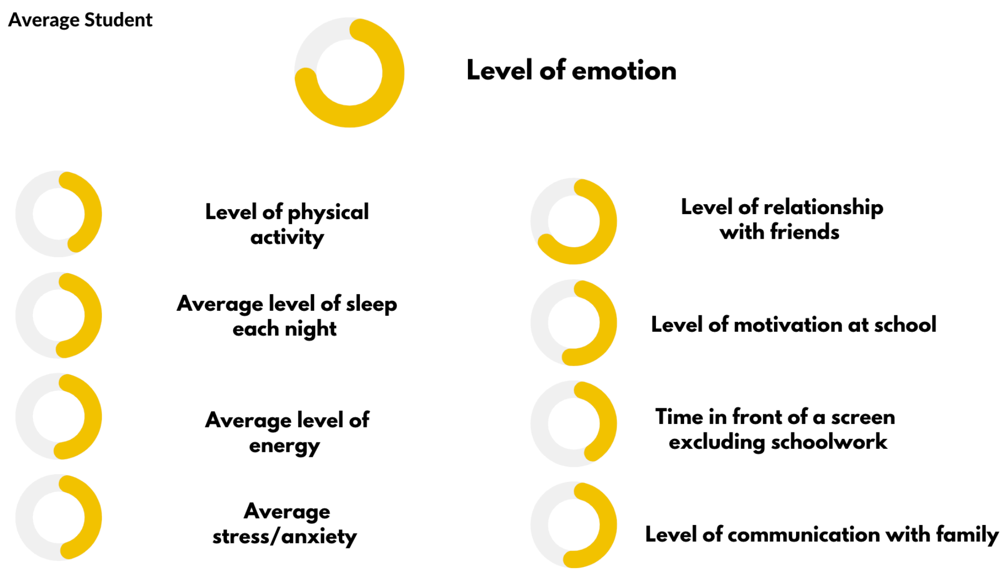

Figure 15: Average Student Cluster

In figure 16, we can describe cluster 2 as being fatigued students. Two significant values that differentiate them from the other students are q4 (energy) and q7 (motivation), in which they both scored low points. The fatigued student cluster scored the lowest in q1 (emotion), q2 (physical activity), q3 (sleep), q6 (relationship with friends), and q9 (communication with family), meaning lowest level of emotion, did the least physical activity, slept the least, had worse relationships with friends and had worst communication with family than the other clusters. This cluster also scored the highest in q5 and q8, meaning high stress/anxiety and high screen time after schoolwork.

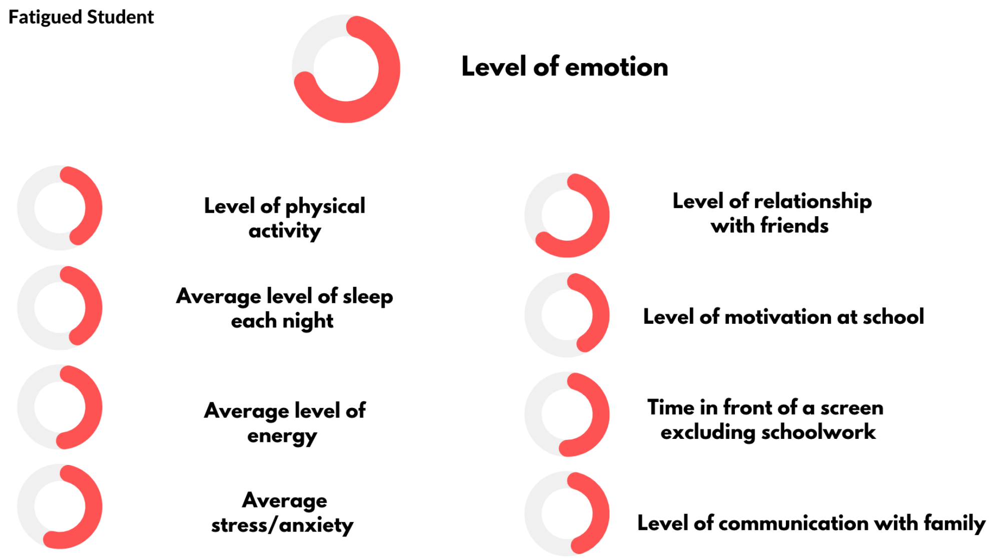

Figure 16: Fatigued Student Cluster

In figure 17, we can describe cluster 3 as energetic students. Two significant values that differentiate them from the other students are energy and motivation, in which they both scored the highest points. The same variables identified as being low in cluster 2 were the highest in cluster 3, meaning they had the highest level of emotion, did the most physical activity, slept the most, had better relationships with friends and had good communication with family. This cluster also scored the lowest in q5 and q8, meaning low stress/anxiety and low screen time after schoolwork.

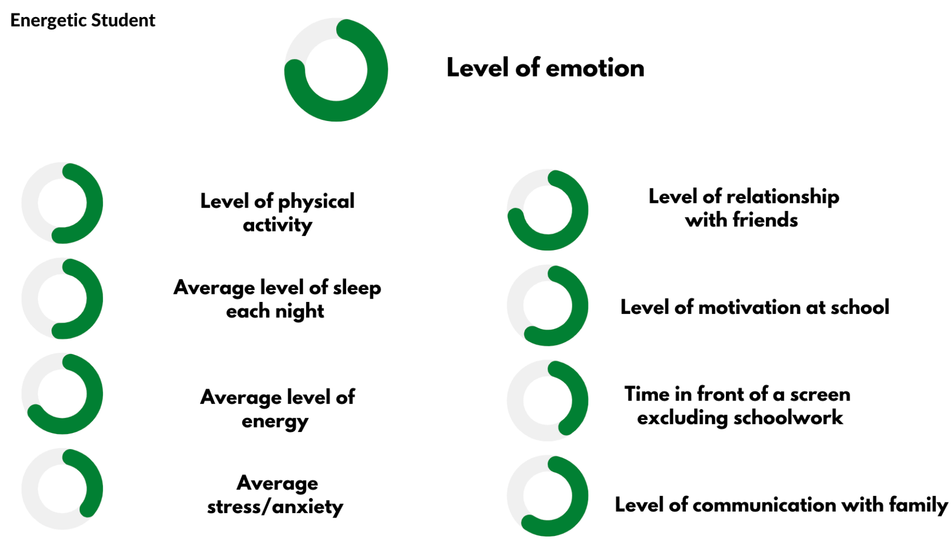

Figure 17: Energetic Student Cluster

In conclusion, we clustered participants based on the evolution of emotion using curve clustering techniques. We identified characteristics that, on average, significantly differ by cluster. We then created personas based on the means of these characteristics by cluster and then we formally compared the means of the characteristics between personas and visualized the differences.

We can safely conclude that we have identified 3 clusters in this study with 3 different personas. The average student, the fatigued students, and the energetic students. In this case, we can clearly identify which group would need help from healthcare professionals in the school setting.

In figure 18, we have a summary of figures 15, 16, and 17 which summarizes by cluster the level of each question.

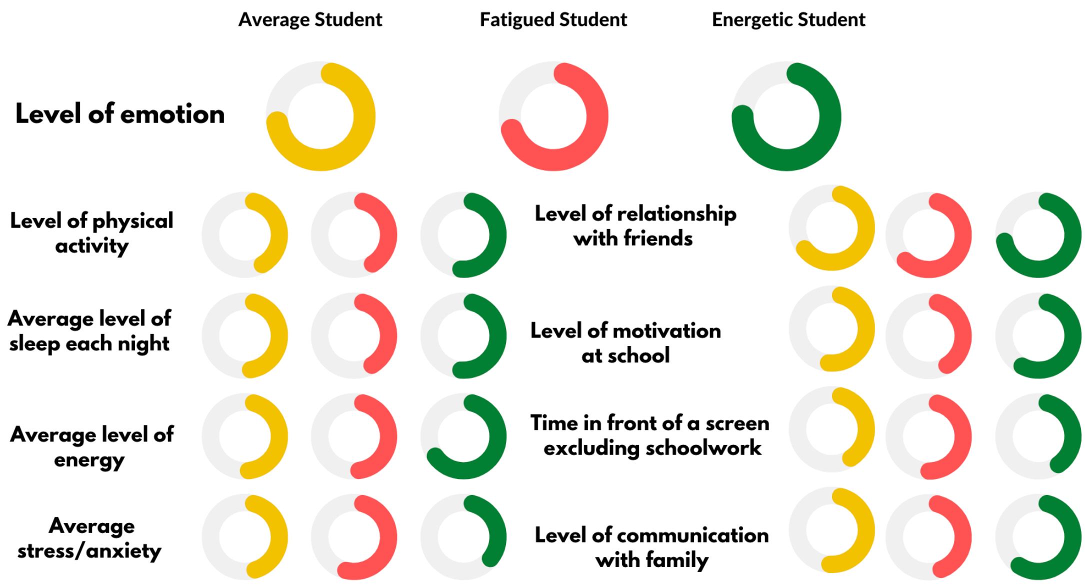

Figure 18: Summary of all clusters

Conclusion

The purpose of this thesis is to find if there are commonalities and/or differences between groups of students. More importantly, we were interested in knowing to what extent we can use curve clustering to find meaningful characteristics in groups of students relative to their mental health. Our analysis would help us understand what influences these specific students to be happier or more anxious based on their weekly activities.

This dissertation has added to our knowledge of the relationship between energy and motivation which affected the clustering of the students. To answer our research questions, Dis-Moi, a non-profit organization, sent out mental health surveys to all grades in seven different high schools. A total of 85 students were part of the final study.

This chapter presents a review of the research questions with related findings, as well as its research limitations, implications, and avenues for future research.

Return on the results

The results of this dissertation addressed the following research questions:
Is it possible to find clusters of students over time based on the longitudinal similarities/differences in their emotions and find characteristics that differentiate these groups?

Our hypothesis is:
With curve clustering analysis techniques over time, we anticipate finding different meaningful clusters of students. With further analysis, we will be able to find which main characteristics differentiating these groups.

Our results revealed that we found three different clusters to describe the students: the average student, the fatigued student, and the energetic student. The results of our study confirmed our hypothesis: there are clusters to be found based on our dataset. We could detect groups of students that could be clustered based on their level of emotion every week.

Furthermore, we used curve clustering to find meaningful characteristics in groups of students relative to their mental health. The results of our study partially support this hypothesis: the curves can be clustered but with this dataset, the clusters lack a clear pattern of trend. In this case, the data is to blame, and the validity of the procedure should not be affected. We can see that in figure 11, all clusters were very close to one another and were crossing at certain weeks.

In addition, our results revealed that two main variables influenced the students to fall in their respective clusters, and those were: energy and motivation. With those clusters found, we could then conclude the cluster with highly energetic and motivated students did better off than any other cluster. And vice versa, low energy, and less motivated students, the fatigued students, were less happy and more anxious as well.

We can bring suggestions to increase students' well-being if they are placed in the "fatigued student" cluster, health workers should focus on their level of energy and motivation in school to help them combat their low level of emotions. We can also suggest that even if the average students aren't at risk, they should still be monitored and given some support to help them succeed and gain happiness along the way.

Research limitations

The results of this dissertation must be interpreted considering its limitations. First, since the students were not forced to answer the survey every week, a lot of data was not usable for this study for lack of data.

Second, to do this study, we assumed that the distance between emotional states is treated as equal everywhere. We also assumed that the scale of emotions was based on Russell's Circumplex Model. In reality, the distance between two emotions could be a lot bigger or smaller than one point. This could have affected our study if we had ordered them any other way.

Third, the correlation-based dissimilarity method had its limitations as it requires a time-lock pattern, meaning students had to answer the survey on the same day every time. In our case, this is a limitation as the method works perfectly, except that the surveys weren't answered on the same dates. This is not the most tolerant method as it isn't flexible with cases having off-sets answers in survey dates. Another method could have been more tolerant and flexible with the data set we had to work with.

Fourth, since this study was self-assessed, the reality of the students' emotions versus the answer they put down in the survey could affect the validity of the research as we cannot confirm if this was effectively their true emotions. This study could not be compared to a clinic diagnostic where a professional can justify their findings. It should also be noted that the fourth cluster was inconclusive because nothing was differentiating it from the other cluster. This cluster might be students who answered the same things every week out of a lack of care for the actual study.

Finally, our cluster lines were very close together. This is typically not the most ideal dataset when clustering as there is a lack of a clear pattern of trend. As mentioned earlier, the data is to blame, and the validity of the procedure should not be affected.

Future research

In future research, it would be interesting to replicate our experiment while asking the same set of questions every week and making sure all students answered the survey. It would be ideal if there was a box for justifying their answers as they can be perceived differently from person to person. On a larger scale, this would help understand the deeper understanding of why they put the scores they did at every question. Furthermore, with the results, school health workers could already have enough data to help students during the study and see if there are improvements instead of waiting until the end to see the results. In our study, the scale was different for quite a few questions making it difficult to evaluate how important those questions truly were to the study, as they were asked different questions every two weeks except the level of emotion making it difficult to see a real trend.

Future studies could also optimize the method used in the present study to increase its validity and adapt it to different contexts. For example, it could be interesting to make the experience more immersive by having students answer this survey after a meditation session in class or after health classes every week.

In conclusion, more research will be needed for a more thorough understanding of how students' emotions should be monitored by health professionals. Creating personas is the first step to understanding the categories of students that are exist, and the actions to take to help each one of them.

In addition, mental health services for minors must confront the additional mental health issues that the COVID pandemic has brought. Managing these repercussions is even more dubious in the absence of recommendations tailored to the new conditions or limited to the worst cases.

Bibliography

Akmal, M., Hares, S., & O'Donnell, M. (2020). Gendered impacts of COVID-19 school closures: Insights from frontline organizations. *Center for Global Development: Washington, DC, USA*.

Aqeel, M., Abbas, J., Shuja, K.H., Rehna, T., Ziapour, A., Yousaf, I. and Karamat, T. (2021). The influence of illness perception, anxiety and depression disorders on students mental health during COVID-19

outbreak in Pakistan: a Web-based cross-sectional survey. *International Journal of Human Rights in Healthcare*, *15*(1), 17-30. https://doi.org/https://doi.org/10.1108/IJHRH-10-2020-0095 (20 August 2021)

Badawi, E., Coursaris, C. K., Sénécal, S., & Léger, P.-M. (2022). User Experience of Universal School-Based e-Mental Health Solutions. International Conference on Human-Computer Interaction,

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of medical Internet research*, *21*(9), e14567.

Benningfield, M. M., Potter, M. P., & Bostic, J. Q. (2015). Educational impacts of the social and emotional brain. *Child and Adolescent Psychiatric Clinics*, *24*(2), 261-275.

Blakemore, S.-J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology*, *65*, 187-207.

Brandmaier, A. M. (2015). pdc: An R package for complexity-based clustering of time series. *Journal of Statistical Software*, *67*, 1-23.

Canada, S. (2019). Health characteristics, annual estimates. https://doi.org/https://doi.org/10.25318/1310009601-eng.

Cheam, A. S., & Fredette, M. (2020). On the importance of similarity characteristics of curve clustering and its applications. *Pattern Recognition Letters*, *135*, 360-367.

Chester, K. L., Callaghan, M., Cosma, A., Donnelly, P., Craig, W., Walsh, S., & Molcho, M. (2015). Cross-national time trends in bullying victimization in 33 countries among children aged 11, 13 and 15 from 2002 to 2010. *The European Journal of Public Health*, *25*(suppl_2), 61-64.

Collishaw, S., Maughan, B., Goodman, R., & Pickles, A. (2004). Time trends in adolescent mental health. *Journal of Child Psychology and psychiatry*, *45*(8), 1350-1362.

Cosma, A., Stevens, G., Martin, G., Duinhof, E. L., Walsh, S. D., Garcia-Moya, I., Költő, A., Gobina, I., Canale, N., & Catunda, C. (2020). Cross-national time trends in adolescent mental well-being from 2002 to 2018 and the explanatory role of schoolwork pressure. *Journal of Adolescent Health*, *66*(6), S50-S58.

Douek, E. (2020). COVID-19 and Social Media Content Moderation. https://www.lawfareblog.com/covid-19-and-social-media-content-moderation

El-Habil, A. M. (2012). An application on multinomial logistic regression model. *Pakistan journal of statistics and operation research*, 271-291.

Figueroa, C. A., & Aguilera, A. (2020). The need for a mental health technology revolution in the COVID-19 pandemic. *Frontiers in Psychiatry*, *11*, 523.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287-298.

Gallant, N., Vachon, N., Sirois-Gaudreau, P., Cozic-Fournier, Y., & Labrecque, K. (2019). Portrait du bien-être des jeunes au Québec: Ensemble du Québec. In: Québec: INRS et Fondation Jeunes en tête.

Golay, X., Kollias, S., Stoll, G., Meier, D., Valavanis, A., & Boesiger, P. (1998). A new correlation-based fuzzy logic clustering algorithm for FMRI. *Magnetic resonance in medicine*, *40*(2), 249-260.

Gutman, L., & Feinstein, L. (2008). *Children's well-being in primary school: Pupil and school effects [Wider Benefits of Learning Research Report No. 25]*. Centre for Research on the Wider Benefits of Learning, Institute of ….

Gysbers, N. C. (2012). Embrace the past, welcome the future: A brief history of school counseling. *American School Counselor Association,(Ed.).(2012). The ASCA national model: A framework for school counseling programs. Alexandria, VA: Author*.

Hanko, G. (2018). *Special needs in ordinary classrooms: From staff support to staff development*. Routledge.

Kalpakis, K., Gada, D., & Puttagunta, V. (2001). Distance measures for effective clustering of ARIMA time-series. Proceedings 2001 IEEE international conference on data mining, Kay, H. (1999). Bright Futures: Promoting children and young people's mental health. *London: The Mental Health Foundation*.

Kessel Schneider, S., O'Donnell, L., & Smith, E. (2015). Trends in cyberbullying and school bullying victimization in a regional census of high school students, 2006-2012. *Journal of school health*, *85*(9), 611-620.

Kutcher, S., Bagnell, A., & Wei, Y. (2015). Mental health literacy in secondary schools: a Canadian approach. *Child and Adolescent Psychiatric Clinics*, *24*(2), 233-244.

Larsen, M. E., Huckvale, K., Nicholas, J., Torous, J., Birrell, L., Li, E., & Reda, B. (2019). Using science to sell apps: evaluation of mental health app store quality claims. *NPJ digital medicine*, *2*(1), 1-6.

Lempinen, L., Junttila, N., & Sourander, A. (2018). Loneliness and friendships among eight-year-old children: Time-trends over a 24-year period. *Journal of Child Psychology and psychiatry*, *59*(2), 171-179.

Liang, L., Ren, H., Cao, R. & al. . (2020). The Effect of COVID-19 on Youth Mental Health. *Psychiatric Quarterly*, *91*, 841–852. https://doi.org/https://doi.org/10.1007/s11126-020-09744-3

Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, *38*(11), 1857-1874.

Liu, S., Yang, L., Zhang, C., Xiang, Y.-T., Liu, Z., Hu, S., & Zhang, B. (2020). Online mental health services in China during the COVID-19 outbreak. *The Lancet Psychiatry*, *7*(4), e17-e18.

McCarthy, B. (2021). Using conversational AI to support employees' psychological needs. https://www.ibm.com/blogs/watson/2021/11/conversational-ai-employee-care/

McMartin, S. E., Kingsbury, M., Dykxhoorn, J., & Colman, I. (2014). Time trends in symptoms of mental illness in children and adolescents in Canada. *Cmaj*, *186*(18), E672-E678.

Montero, P., & Vilar, J. A. (2014). TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software*, *62*(1), 1 - 43. https://doi.org/10.18637/jss.v062.i01

Moussa, Y., Mahdanian, A. A., Yu, C., Segal, M., Looper, K. J., Vahia, I. V., & Rej, S. (2017). Mobile health technology in late-life mental illness: a focused literature review. *The American Journal of Geriatric Psychiatry*, *25*(8), 865-872.

Naslund, J. A., Aschbrenner, K. A., McHugo, G. J., Unützer, J., Marsch, L. A., & Bartels, S. J. (2019). Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. *Early intervention in psychiatry*, *13*(3), 405-413.

Nielsen, R., Fletcher, R., Newman, N., Brennen, J., & Howard, P. (2020). *Navigating the 'infodemic': How people in six countries access and rate news and information about coronavirus*. Reuters Institute for the Study of Journalism.

Organization, W. H. (2020). Guidelines on mental health promotive and preventive interventions for adolescents: helping adolescents thrive.

Organization, W. H. (2021). Mental health in schools: a manual. *Regional Office for the Eastern Mediterranean*, 98. https://apps.who.int/iris/handle/10665/347512

Patalay, P., & Gage, S. H. (2019). Changes in millennial adolescent mental health and health-related behaviours over 10 years: a population cohort comparison study. *International journal of epidemiology*, *48*(5), 1650-1664.

Power, J. (2019). US telehealth satisfaction study. *Agoura Hills, CA: SM, JD Power*.

Primack, B. A., Shensa, A., Escobar-Viera, C. G., Barrett, E. L., Sidani, J. E., Colditz, J. B., & James, A. E. (2017). Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among US young adults. *Computers in human behavior*, *69*, 1-9.

Ralston, A. L., Andrews III, A. R., & Hope, D. A. (2019). Fulfilling the promise of mental health technology to reduce public health disparities: Review and research agenda. *Clinical Psychology: Science and Practice*, *26*(1), 60.

Ramsay, J. O., & Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 539-561.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Rutter, M. (1991). Pathways from childhood to adult Life; the role of schooling. *Pastoral Care in Education*, *9*(3), 3-10.

Sadler, K., Vizard, T., Ford, T., Marchesell, F., Pearce, N., Mandalia, D., Davis, J., Brodie, E., Forbes, N., & Goodman, A. (2018). Mental health of children and young people in England, 2017.

Services, U. S. D. o. H. H. (2022). How much physical activity do children need? *Center of Disease Control and Prevention*. https://www.cdc.gov/physicalactivity/basics/children/index.htm#print

Silver, L. (2019). Smartphone ownership is growing rapidly around the world, but not always equally.

Simpson, K. R., Meadows, G. N., Frances, A. J., & Patten, S. B. (2012). Is mental health in the Canadian population changing over time? *The Canadian Journal of Psychiatry*, *57*(5), 324-331.

Stevenson, D. L., & Baker, D. P. (1987). The family-school relation and the child's school performance. *Child development*, 1348-1357.

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school-and individual-level factors on academic achievement. *Education and urban society*, *40*(2), 179-204.

Street, M. (2018). Environnement social: la famille, les amis, l'école et la communauté. *Enquête québécoise sur la santé des jeunes du secondaire 2016-2017*, 31-65.

Sweeting, H., West, P., Young, R., & Der, G. (2010). Can we explain increases in young people's psychological distress over time? *Social science & medicine*, *71*(10), 1819-1830.

Team, C. (2022). Correlation Matrix. https://corporatefinanceinstitute.com/resources/excel/study/correlation-matrix/

Tully, E. C., Iacono, W. G., & McGue, M. (2008). An adoption study of parental depression as an environmental liability for adolescent depression and childhood disruptive disorders. *american Journal of Psychiatry*, *165*(9), 1148-1154.

Waite, P., Pearcey, S., Shum, A., Raw, J. A., Patalay, P., & Creswell, C. (2021). How did the mental health symptoms of children and adolescents change over early lockdown during the COVID-19 pandemic in the UK? *JCPP advances*, *1*(1), e12009.

Xie, X., Xue, Q., Zhou, Y., Zhu, K., Liu, Q., Zhang, J., & Song, R. (2020). Mental health status among children in home confinement during the coronavirus disease 2019 outbreak in Hubei Province, China. *JAMA pediatrics*, *174*(9), 898-900.

Zwaanswijk, M., van Dijk, C. E., & Verheij, R. A. (2011). Child and adolescent mental health care in Dutch general practice: time trend analyses. *BMC family practice*, *12*(1), 1-11.

Annex

Annex 1: R Code

```
#install packages and libraries

#install.packages('plyr', repos = "http://cran.us.r-project.org")
#install.packages("rgl")
library(rgl)
#install.packages("TSclust")
library(TSclust)
#install.packages("sqldf")
library(sqldf)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("directlabels")
library(directlabels)
#install.packages("readxl")
library(readxl)
#install.packages("xlsx")
library(xlsx)

#dataSet
cleanData <- read_excel("/Users/leojes00/Desktop/ImputeMostFrequentTransposed.xlsx")
View(cleanData)

#clustering
dis="COR"
d<-diss(cleanData[,-1],dis)

#choose the N clusters and convert cl to dataFrame then convert to excel/csv
n2<-kmeans(d,2)$cluster
```

```
x <- data.frame(as.list(n2))
x <- t(x)
write.xlsx(x, "/Users/leojes00/Desktop/Q1RN2.xlsx", sheetName="Sheet1")

n3<-kmeans(d,3)$cluster
x <- data.frame(as.list(n3))
x <- t(x)
colnames(x)[1]  <- "Clusters"
write.xlsx(x, "/Users/leojes00/Desktop/Q1RN3.xlsx", sheetName="Sheet1")

n4<-kmeans(d,4)$cluster
x <- data.frame(as.list(n4))
x <- t(x)
colnames(x)[1]  <- "Clusters"
write.xlsx(x, "/Users/leojes00/Desktop/Q1RN4.xlsx", sheetName="Sheet1")
```